

UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK

-----		X
H. CRISTINA CHEN-OSTER; LISA PARISI; and	:	
SHANNA ORLICH,	:	
	:	
Plaintiffs,	:	
	:	
v.	:	10 Civ. 6950 (AT) (JCF)
	:	
GOLDMAN, SACHS & CO. and THE GOLDMAN	:	
SACHS GROUP, INC.	:	
	:	
Defendants.	:	
-----		X

**EXPERT REPORT OF MICHAEL P. WARD**

**Part 1 of 2**

**Expert Report of Michael P. Ward, Ph.D.**  
**in the Matter of**  
***Chen-Oster et al. vs.***  
***Goldman, Sachs & Co. and The Goldman Sachs Group, Inc.***  
**July 3, 2014**

## TABLE OF CONTENTS

	Page
INTRODUCTION.....	1
BACKGROUND AND QUALIFICATIONS.....	1
SUMMARY .....	2
PART I – DR. FARBER’S STUDIES ARE FUNDAMENTALLY FLAWED.....	2
DR. FARBER’S STUDIES DO NOT SHOW A COMMON PATTERN OF ADVERSE TREATMENT .....	5
I.    DR. FARBER’S PAY STUDIES CONFIRM THAT THERE IS NO COMMON PATTERN OF ADVERSE TREATMENT .....	5
II.   THERE IS NO COMMON PATTERN ADVERSE TO WOMEN IN 360 REVIEW SCORES .....	7
III.  THERE IS NO COMMON PATTERN ADVERSE TO WOMEN IN MANAGER QUARTILE ASSIGNMENT .....	7
IV.   DR. FARBER’S MODELS ERRONEOUSLY AGGREGATE LARGE GROUPS OF EMPLOYEES INTO A SINGLE STATISTICAL MODEL .....	8
V.    DR. FARBER USES INCORRECT VALUES FOR COMPENSATION.....	12
VI.   DR. FARBER’S STUDIES FAIL TO MODEL THE PROCESSES HE STUDIED .....	13
A.    Dr. Farber’s Use of Proxies for Productivity and Job Function in His 360 Review Process Regression Models Was Improper .....	13
B.    Dr. Farber’s Manager Quartile Regression Models Are Flawed .....	14
C.    When the Flaws in Dr. Farber’s Revised Pay Regression Models Are Corrected, the Gender Differences He Reports Decrease Dramatically .....	15
D.    Dr. Farber’s Models Do a Very Poor Job of Explaining Pay Differences Within the Groups He Studies.....	20
VII.  DR. FARBER’S GENDER DIFFERENCE IS NOT COMPELLING GIVEN HOW MUCH VARIATION IN PAY REMAINS UNEXPLAINED .....	21
VIII. DR. FARBER’S MODEL PREDICTS SIMILAR PAY FOR EMPLOYEES WHOSE ACTUAL PAY IS VERY DIFFERENT .....	22
Some Examples of Dr. Farber’s “Unexplained” Pay Differences.....	24
IX.   DR. FARBER’S CONCLUSIONS ABOUT THE IMPACT OF MANAGER QUARTILE AND 360 REVIEW PROCESSES ON GENDER DIFFERENCES IN PAY ARE OVERSTATED AND NOT WELL-FOUNDED .....	25
A.    The Impact of Manager Quartile and 360 Reviews on Pay Varies Widely Across the Putative Class .....	26
B.    Dr. Farber Overstates the Importance of Manager Quartile Assignment and 360 Review Scores on Gender Differences in Pay.....	27
X.    DR. FARBER’S PROMOTION STUDIES ARE INCONSISTENT WITH HIS OTHER STUDIES; WHEN PERFORMED CONSISTENTLY WITH HIS OTHER MODELS, THE RESULTS ARE NEUTRAL TO WOMEN .....	28

## TABLE OF CONTENTS

	Page
PART II – MY STUDIES AND ANALYSES .....	30
XI. THE STRUCTURE OF THE FIRM AND THE PROCESS BY WHICH COMPENSATION DECISIONS ARE MADE AT GOLDMAN SACHS MUST SHAPE ANY PROPER STATISTICAL ANALYSIS.....	31
XII. PLAINTIFFS' CRITICISMS OF THE FIRM'S MANAGER QUARTILE ASSESSMENTS ARE NOT SUPPORTED BY THE DATA.....	33
A. Manager Quartile Assignments are Significantly Positively Related to 360 Reviews.....	33
B. Manager Quartile Assignments are Significantly Related to "Production Oriented" Measures of Performance As Well .....	36
C. The Manager Quartile Process is Gender Neutral.....	38
XIII. GOLDMAN SACHS DATA SHOW WIDE VARIATION IN PAY .....	41
XIV. MY REGRESSION ANALYSIS PRESENTS A FULLER AND MORE ACCURATE ANALYSIS OF COMPENSATION .....	44
A. "Similarly Situated" Employees .....	44
B. Similar Business Conditions .....	45
C. Similar Work.....	46
D. Similar Performance.....	47
E. Many Unmeasurable Factors and Approximate Values Affect Variation in Compensation .....	48
F. My Regression Models Better Reflect These Concepts Than Dr. Farber's.....	52
G. Lateral Hires.....	53
XV. ANALYSIS OF COMPENSATION .....	55
A. There is No Pattern of Pay Differences across Business Units.....	55
B. Over All Divisions, Business Unit Statistics Do Not Show a Pattern of Compensation Differences.....	60
C. Statistical Findings .....	61
1. Securities .....	62
2. Investment Banking Division.....	65
3. Investment Management Division.....	66
4. Merchant Banking Division .....	68
5. Laterals .....	69
XVI. ROUND CHANGES .....	71
A. PATC Changes Over Compensation "Rounds" Advantage Women More Than Men .....	71
XVII. ANALYSIS OF PROMOTION OF VICE PRESIDENTS.....	73
A. Selection Processes.....	73
B. Statistical Analysis .....	74

## INTRODUCTION

In this Report, I re-state the opinions previously set forth in my Reports dated December 13, 2013 and May 7, 2014.<sup>1</sup>

## BACKGROUND AND QUALIFICATIONS

I am Vice President and Senior Analyst at Welch Consulting, a consulting firm specializing in economic and statistical research. I am also senior economist at Unicon Research Corporation, a sister organization performing grant and contract research for the Federal Government. I am a member of the American Economic Association, the American Statistical Association, the American Compensation Association (World at Work). Attached hereto as Appendix B is a copy of my resume.

I hold a bachelor's degree in Economics from the University of California, Santa Barbara, and M.A. and Ph.D. degrees in Economics from the University of Chicago. My fields of specialty in graduate school were labor economics and econometrics. Econometrics is the application of statistical methods to economic data. From 1973 to 1975, I taught economics and statistics at the graduate and undergraduate level at the University of California, Santa Barbara. From 1975 to 1980, I taught economics and statistics at the graduate and undergraduate level at the University of California, Los Angeles. From 1980 to 1984, I was employed full-time as Senior Economist at the Rand Corporation. In 1984, I joined Welch Consulting and Unicon Research Corporation. In 1983, I was appointed by the U.S. Secretary of Commerce as a member of the U.S. Census Advisory Committee on Population Statistics, and I remained a member of that Committee until 1988. For two years, I was chairman of that committee.

My research, conducted while I was at UCLA, Rand, and at Unicon, was supported by the Department of Labor, the National Institutes of Health, the National Institute of Aging, the Department of the Air Force, the Department of Energy, and the Assistant Secretary of Defense. Virtually all of that work was in the area of labor economics and statistics. The product of this research is shown in my resume, attached as Appendix B.

Over the past 30 years, in my capacity as a consulting economist, I have frequently been retained by major corporations to perform statistical studies of employment practices. I have testified in Federal and State Courts as an expert witness on issues relating to both economics and statistics. I have conducted, and have testified about, statistical studies of employment practices, including hiring, pay, promotion, and terminations. Attached as Appendix C is a list of all cases in which I have testified in court, or in which I have given deposition testimony from 2009 to present. My billing rate is \$650 per hour.

---

<sup>1</sup> I submitted the Expert Report of Michael P. Ward, Ph.D on December 13, 2013 and the Expert Sur-Rebuttal Report of Michael P. Ward, Ph.D on May 7, 2014.

**SUMMARY**

I have been asked by counsel for Goldman Sachs to perform two separate tasks related to the discrimination claims raised by Plaintiffs in this putative class action. First, I was asked to critique the work of Plaintiffs' statistical expert, Dr. Henry Farber,<sup>2</sup> who studied compensation, performance, and promotion decisions in three of Goldman Sachs' Revenue Divisions.<sup>3</sup>

Second, I was asked to analyze the data myself and offer my own expert opinions on the allegations that discrimination in pay, promotion, and performance review is common to the putative class alleged in the First Amended Complaint. My critique of Dr. Farber's work is set forth in Part I below, and in Part II, I describe the results of my own analysis and the methodology I used.

**PART I – DR. FARBER'S STUDIES ARE FUNDAMENTALLY FLAWED**

The studies in Dr. Farber's Report do not support the opinions he offers for several fundamental reasons. I address them below.

**Lack of Common Adverse Effect**

Contrary to his stated conclusions, Dr. Farber's studies confirm that the personnel processes Plaintiffs challenge have *no* common effect on women in the putative class. While Dr. Farber presents aggregated average outcomes for the entire putative class, his studies are erroneously based on the assumption—which Dr. Farber never substantiates—that the gender differences he claims to find are consistently adverse to women across all of the many businesses and groups he studies. As I explain below in Sections I, II, and III, the data show that Dr. Farber's assumption is demonstrably incorrect.

Even using the flawed methods in Dr. Farber's Report, when his results are displayed and reviewed at the Business Unit level—the level at which relevant decisions are made—women do *better* than men with respect to 360 reviews and manager quartile assignment in more than 30% of the Business Units in the population Dr. Farber studies, and women do *better* than men with respect to pay in more than 40% of those Business Units. Overall, in 75% of all the Business Units studied by Dr. Farber, the pay differences are gender neutral (in other words, within the range that would be expected by chance). These percentages are all based on Dr. Farber's own models.

---

<sup>2</sup> Dr. Farber submitted four reports in this case, an Initial Report dated October 30, 2013, an Addendum Report dated December 3, 2013, a Rebuttal Report dated January 28, 2014, and then a February 17, 2014 Report, which was filed with Plaintiffs' Motion for Class Certification. Unless otherwise identified herein, I use "Report" to refer to Dr. Farber's February 17, 2014 Report, which was filed with Plaintiffs' Motion for Class Certification.

<sup>3</sup> Dr. Farber initially studied four Divisions – Securities, the Investment Management Division ("IMD"), the Investment Banking Division ("IBD"), and the Merchant Banking Division ("MBD"). However, in his Rebuttal Report and in the Report that was filed with Plaintiffs' Motion for Class Certification, Dr. Farber inexplicably dropped MBD from his studies. Because Dr. Farber initially studied MBD, I studied that Division as well, and several of those studies are reflected in Part II of this Report.

Dr. Farber's own models show no common adverse effect on women resulting from the 360 review, manager quartile assignment, or pay processes at Goldman Sachs.

### **Failure to Apply Standard Statistical Techniques**

Not only do Dr. Farber's models fail to show a common pattern in any of the processes he is studying, those models are weak and unreliable because they violate standard statistical methods. As I describe in Section IV below, to make meaningful comparisons, a proper statistical study must compare similarly situated employees to one another. Dr. Farber, however, ignores this bedrock principle and combines into a single regression model employees hired laterally and non-laterals, and employees in all of the Goldman Sachs Divisions he studies in his Report. He does so without performing the standard statistical test (called the "Chow" test) that tells statisticians whether or not the models for such groups are sufficiently homogeneous to be analyzed together. I perform the Chow test, and the results show that employees across different Divisions cannot be combined in a single pay model. Nor can employees who were hired laterally be combined with those who were promoted from within the firm. The Chow test rejects combining these populations in one statistical model because of the differential impact of the factors in the model on the pay of each group's members. Dr. Farber's work in his Report fails this standard test.

### **Correcting Fundamental Flaws in Dr. Farber's Model**

As described in Section VI below, the statistical gender differences Dr. Farber claims to have found rest largely on factors and approaches that bear no relationship to the Goldman Sachs processes he is studying. His models are fundamentally defective. When those flaws are corrected, the pay results for 85% of class members studied are gender-neutral with outcomes that are not statistically significantly adverse to women. More particularly:

- In addition to combining disparate groups that require separate examination under standard statistical techniques, Dr. Farber refuses to account for critical factors in Goldman Sachs' processes—such as the Business Units in which employees work and measures of employee productivity and performance—and he insists on using extraneous factors (such as Affirmative Action Plan ("AAP") job groups), which are not inputs relevant to performance and compensation decision-making at Goldman Sachs.<sup>4</sup> He also continues to use irrelevant "proxy" variables for job function and productivity while refusing to use the actual Goldman Sachs data available to him.

---

<sup>4</sup> Dr. Farber also includes employees who are not in the putative class in his studies. For example, he includes non-revenue employees in his models, along with Private Wealth Advisors, commission-based employees paid on a different pay structure than other putative class members and who I have been told signed arbitration agreements with Goldman Sachs.

- As a result of these glaring flaws in his analysis, Dr. Farber's results explain only a very small part of the variation in pay he observes. Quite apart from any claimed gender differences, his models do a poor job of explaining variations in pay even when applied only to men. This tells me that Dr. Farber is not accounting for the factors that actually impact pay at Goldman Sachs. This is apparent from the "R-squared" of his models, which is the measure most commonly used by statisticians to quantify the explanatory power of a statistical model. Dr. Farber does not report the "R-squared" for his studies. The backup data to his studies, however, show a very low or poor "R-squared" for his models.
- Correcting for the serious flaws in Dr. Farber's methodology dramatically shrinks the gender differences Dr. Farber claims to have identified and just as dramatically improves the model's R-squared. When Dr. Farber's studies are revised so that only similarly situated employees are compared to one another, and the major factors *indisputably* considered in setting compensation at Goldman Sachs are taken into account, the R-squared (or explanatory power) of the model *more than doubles*,<sup>5</sup> and, as noted above, there are no statistically significant differences adverse to women in pay for *85% of class members analyzed*.
- As a result of the flaws in his model, Dr. Farber also overstates the impact of the manager quartile and 360 processes on gender differences in pay in his studies. The manager quartile and 360 processes are highly correlated with pay generally (for both men and women), but they do not have the impact on reported gender pay differences that Dr. Farber attributes to them. This is because statistical studies (both mine and Dr. Farber's (when displayed properly)) show that the differences between how men and women fare in the 360 review and manager quartile processes vary widely, and the results are infrequently statistically significant. Given the small and varied discrepancies between men and women in these processes, the impact of these discrepancies in explaining pay differences between men and women is similarly small and varied. In other words, these processes do not have the large uniform impact on gender pay differences that Dr. Farber attributes to them.

### Promotions

Finally, in Section X, I show that Dr. Farber's method of studying promotions is inconsistent both with the approach he endorses earlier in his report and with how promotions actually work at Goldman Sachs. If Dr. Farber had simply followed the approach he applied to his other studies, the results of his promotion study would have not been statistically adverse to women.

---

<sup>5</sup> The R-squared of the corrected models is more than double the R-squared of Dr. Farber's preferred model specification, which does not include controls for any performance measures.



**DR. FARBER'S STUDIES DO NOT SHOW A COMMON PATTERN OF ADVERSE TREATMENT**

None of the studies or models in Dr. Farber's Report examines (or was designed to examine) whether there is a common pattern of treatment of women across the putative class. When the results Dr. Farber obtained are examined by the Business Units in which the members of the putative class work—which is the level at which compensation budgets are set and compensation decisions are made—I see the following results:

**Pay differences:** 42% of the Business Units are *favorable* to women;<sup>6</sup>

**360 scores:** 32% of the Business Units are *favorable* to women; and

**Manager quartile assignment:** 32% of the Business Units are *favorable* to women.

And while there are some Business Units in which there are differences adverse to women, in the overwhelming majority (**75%**) of Business Units, the differences are statistically insignificant. In the charts attached in Appendix A to this report, I show Dr. Farber's own statistical results and highlight the variety of outcomes reflected in each study. Each statistical result is "color coded" to show whether the result favors or disfavors women, and whether or not that result is statistically significant. The visual display shown in these charts (as well as the statistics that underlie them) reveal, in fact, no patterns adverse to women. If such adverse patterns existed or if the pay or performance processes commonly worked to the detriment of women, I would expect to see a significant percentage of the analyses yield adverse outcomes that were statistically significant (large enough to be distinguishable from random chance). As I show, there are few instances where this is the case, along with other instances where women are statistically significantly *advantaged*.

All of the results described below are derived from Dr. Farber's own studies.

**I. DR. FARBER'S PAY STUDIES CONFIRM THAT THERE IS NO COMMON PATTERN OF ADVERSE TREATMENT**

As discussed below, Dr. Farber's pay studies – when displayed by Business Unit - show no common pattern of adverse treatment of women.

Dr. Farber presents a single *summary* measure—the average pay difference between males and females across the entire group being studied—as the sole support for his conclusions of commonality. But the "average" differences he reports mask wide variations from Business Unit to Business Unit. That is, using Dr. Farber's own, albeit flawed, methodology, but displaying the results by Business Unit, the pay results for women in some Business Units are favorable and in others they are adverse; there is no common pattern, and statistically significant results are infrequent in either direction.

---

<sup>6</sup> This rises to 45% of the Business Units when I also separate lateral hires and non-lateral employees, a distinction that the statistics demonstrate is necessary, as I discuss below.

Reviewing Dr. Farber's results by Business Unit is critical. The Business Unit is the level at which budgets are allocated within the Divisions; employees within these units are paid from and compete for the same pool of money and are managed by the same set of supervisors. Employees in one Business Unit cannot meaningfully be compared to employees in other Business Units, because they are paid from separate budget allocations.

Dr. Farber's models ignore the fact that compensation decisions at Goldman Sachs are made at the Business Unit level. His model anticipates that a man and woman who are otherwise similar but one of whom works in a Business Unit that has had a phenomenally successful year while the other in a Business Unit that has struggled – will be paid the same, even though they are paid from different bonus pools of often dramatically different sizes.

Accordingly, reviewing the results of any statistical study by the set of Business Units within each Division is necessary to determining whether there are common patterns of treatment of women across the class. At this stage of the case, the appropriate question for the statistical expert to attempt to answer is this: In light of Plaintiffs' allegation of a *pattern* of adverse treatment against women, do the data show a pattern of adverse results for women across the relevant decision-making units? Dr. Farber performs no such analysis.

When I do that analysis using Dr. Farber's own flawed statistical models, however, the results show that **over 42% of the Business Units in the population Dr. Farber analyzed show pay differences that favor women**. See Appendix Chart 1.<sup>7</sup> Perhaps even more critically, in terms of addressing the class requirement of commonality, more than 75% of the Business Units relevant to this case<sup>8</sup> show no *statistically significant* differences in *either* direction.<sup>9</sup>

And the fact that 42% of Business Units show results *favorable* to women is not a result based on any complex statistical test that might be affected by a small sample size and therefore is not subject to criticism as some sort of statistical gaming. These results were generated simply by counting "plusses" and "minuses"—pay differences that are

---

<sup>7</sup> All of the charts referenced in this Report (Sections I, II, III, and IX) are attached hereto in Appendix A. Chart 1 contains color-coded pay differences, with dark green indicating a statistically significant difference between the pay of men and women within the Business Unit that is favorable to women. Light green indicates a favorable difference for women that is not statistically significant. Yellow indicates that the difference is adverse to women but not statistically significant, while dark red indicates that the difference is adverse to women and statistically significant. The results reflected in this chart are from Dr. Farber's own aggregate pay model broken out by Division and Business Unit; I do not estimate any new models of pay to calculate the numbers displayed in this chart but retain Dr. Farber's limited set of regression variables.

<sup>8</sup> I restrict this analysis to the time period from 2005 forward because Business Unit is not clearly identified in the data for 2003 and 2004. There are an additional eight Business Units that have too few employees for me to determine statistical significance. In four of them, pay differences are favorable to women, and in the other four pay differences are adverse to women.

<sup>9</sup> Dr. Farber has criticized my method of measuring the statistical significance of these Business Unit level measures of pay difference, yet he offers no method of his own. He has claimed that I have failed to take account of the uncertainty introduced by the use of pay differences estimated for the larger Division-wide group and then displayed by Business Unit. However, this would not change my findings. Indeed, this supposed uncertainty would tend only to decrease the statistical significance of any Business Unit differences adverse to women that I estimate without affecting at all the proportion (42%) of Business Units showing pay differences that favor women. Therefore I do not find his criticism well-taken.

either favorable or unfavorable to women, based on Dr. Farber's own results. The enormous variation—using Dr. Farber's own models—demonstrates the absence of any common pattern of adverse pay results for women.

Had Dr. Farber simply used his own statistical model and simply divided his single pay difference into results by Business Unit, he would have found that even with his simplistic model, gender differences are not common across these units. Pay differences in Business Units sometimes favor women and sometimes disfavor women. Like me, he would have found no support for a conclusion that there are common class-wide pay differences adverse to women.

## II. THERE IS NO COMMON PATTERN ADVERSE TO WOMEN IN 360 REVIEW SCORES

A similar result is obtained using Dr. Farber's Report model to study the 360 review process. (Tables 14 and 15 of the Report). As I did above with his pay results, I take Dr. Farber's 360 regression model and—without correcting his mistakes or changing any of the factors included in his model—I report his results by Division and Business Unit.<sup>10</sup> See Appendix Chart 2. Overall, Dr. Farber's 360 review results are *favorable* to women (meaning that they receive higher 360 review scores than men) 32% of the time, and sometimes to a statistically significant degree, and are statistically significantly *adverse* to women in only 15% of the Business Units.<sup>11</sup>

I also report the raw differences (that is, I do not control for any factors at all) in the 360 review scores between men and women in the population Dr. Farber includes in his studies. I merely display them by Division and by Business Unit. Approximately 34% of the Business Unit-level differences in 360 review scores are *favorable* to women and over 82% are statistically insignificant in either direction. See Appendix Chart 3.<sup>12</sup> In only 13.4% of the Business Units are there statistically significant differences adverse to women.

It is evident from these numbers that Dr. Farber's 360 regression studies show no pattern of 360 review scores adverse to women across the class.

## III. THERE IS NO COMMON PATTERN ADVERSE TO WOMEN IN MANAGER QUARTILE ASSIGNMENT

In his Report, Dr. Farber also includes regression studies purporting to study the manager quartile assignment process. However, he only looks at whether women were placed in the top quartile at the rates one would expect, and, significantly, he does not attempt to show whether there is a common *pattern* of assignment of women into lower manager

---

<sup>10</sup> Again, I restrict this analysis to the time period from 2005 forward because although the Divisions were operating with Business Units during 2003 and 2004, the data for making a clear identification is not available.

<sup>11</sup> There are an additional 18 Business Units that have too few employees for me to determine statistical significance. In 14 of them, 360 scores are favorable to women, and in four of them, 360 score differences are adverse to women.

<sup>12</sup> As above, I restrict this analysis to 2005 forward. Again, there are 18 Business Units for which statistical significance cannot be determined. Among these, the 360 score differences are favorable to women 14 times and unfavorable to them 4 times.

quartiles across the putative class. Both analyses are important when examining whether there are common patterns of lower manager quartile assignment of women than would be expected. His results show there is no such pattern.

To demonstrate that Dr. Farber's Report analysis is incomplete and misleading, I perform a set of studies that incorporate the same set of regression factors used by Dr. Farber and analyze differences in the assignment of men and women into the top quartile, the middle quartile (2/3) and the bottom quartiles (4 and 5).<sup>13</sup> Appendix Chart 4 shows that in 32.1% of the Business Units, Levels (Associate and Vice President), and Divisions studied, the manager quartiles are *higher* for women than men.<sup>14</sup>

I also calculate the raw differences between men and women in assignment to the highest quartile, and in assignment to quartiles 1, 2 and 3, as contrasted to quartiles 4 and 5.<sup>15</sup> These comparisons by Division, level, and Business Unit are shown in Appendix Chart 5. In approximately 36.1% of Business Units, women receive more *favorable* positions in the quartiles relative to men. This finding is not based on measures of statistical significance but simply counts the units that are favorable to women. For 87.2 % of Business Units, the results are statistically insignificant in either direction.

These results show enormous variation that belies the notion of a common adverse impact—a pattern—across Business Units in the quartile assignment for female Associates and Vice Presidents over the period being studied. Dr. Farber's manager quartile regression studies do not show a common pattern of treatment of women, favorable or unfavorable, in the Business Units across the putative class.

Based on Dr. Farber's own studies, I do not see any common pattern of treatment with respect to gender differences in pay, 360 review scores, or manager quartile assignment. The findings above are based on Dr. Farber's own models and results, albeit properly displayed by Business Unit. However, as discussed in detail below, those studies suffer from several other serious flaws which, when corrected, paint an even more varied picture and, to the extent there are any gender discrepancies, they are far less significant than reported by Dr. Farber.

#### **IV. DR. FARBER'S MODELS ERRONEOUSLY AGGREGATE LARGE GROUPS OF EMPLOYEES INTO A SINGLE STATISTICAL MODEL**

As Dr. Farber implicitly acknowledges, regression models can only meaningfully estimate gender differences in pay if they compare similarly situated men and women. For there to be a meaningful comparison of similarly situated men

---

<sup>13</sup> I report these differences in Appendix Chart 4.

<sup>14</sup> Again, I restrict this analysis to the time period from 2005 forward because Business Unit is not clearly identified in the data for 2003 and 2004. There are two additional Business Units that are too small for me to determine statistical significance. Gender differences are favorable in one and unfavorable in the other.

<sup>15</sup> As with the raw 360 studies described above, these studies are based on the population Dr. Farber includes in his studies.

and women, the statistical models comparing them must themselves be similar, that is, they must give similar weight to factors that affect the outcome (pay or performance). To use Dr. Farber's examples, in comparing the pay of male and female librarians, the factors that affect pay should have a similar impact on the members of this group. Similarly, when comparing the pay of male and female engineers, those factors should have a similar impact on the men and women in the engineer group. It would make no sense to put librarians and engineers into a single statistical model and to explain pay differences by "highest degree earned" or "whether hired laterally." When the factors explaining pay or performance have a different impact depending on the work performed, it makes logical—and economic—sense to separate them.

However, Dr. Farber's regression models combine large numbers of very dissimilar employees: employees hired laterally and those promoted internally into their positions; those who work in different Divisions; and those who serve in different functions.

In part, the Divisional separation is necessary because different changes over time in the revenue and profitability of the different Divisions will translate into changes in compensation for the employees in each Division, and these swings will be different for different Divisions in different years. For example, IBD's profitability was on average higher than IMD's from 2005 to 2007, but IMD did dramatically better than IBD in 2009. The results in these Divisions in these years had a very direct impact on the compensation of the employees working in those Divisions, but not on the compensation of employees in other Divisions. The only way to capture this is to do separate studies by Division.

Instead of performing separate statistical models for each of these populations, Dr. Farber attempts to address the issue by adding "controls" for these various factors in his single regression. This is not an adequate substitute. "Controlling" for a factor means that the incremental value of that factor will be averaged over all of the employees in the study to ascertain whether they are paid as the model predicts they should be; in other words, if one Division did very well in one year while another did very poorly, the average dollar amount expected to be earned by the employees in each of the Divisions will be the average of both. Averaging this value over the entire population masks the fact that only some of the sub-groups in the population actually benefited (or suffered) as a result of the financial results in a particular year, and the resulting analysis *by definition* is not accurate for either Divisional sub-group. Thus, using Division as a control is not sufficient; it necessarily treats as equivalents things we know to be different. As a result, Dr. Farber's flawed model produces statistically unreliable results.

The difference between doing separate studies and simply "adding a control" for some factors in a single regression is critically important. If one studies all of these groups in a single model and simply adds a factor like education into the model one necessarily is assuming that the impact of that variable is uniform across all of the sub-groups in the analysis. Thus, the key determinant of how to structure the analysis is whether the explanatory variables

have similar effects on pay across the different groups. For example, the pattern of variation in profitability over time is different for each Division, resulting in Division-specific patterns of compensation fluctuations over the period. The assumption that year-to-year compensation changes are identical for people in different Divisions is no more likely to be valid than for people in different companies in the same industry. These distinctive patterns mean that, at the very least, the effect of Division on compensation should be allowed to vary over time. This could be accomplished by creating separate models for each Division and controlling for year. Dr. Farber's aggregate methodology assumes that the profitability of all Divisions increases or decreases from one year to the next in lock-step. The data used by Dr. Farber in his regression analysis shows this assumption to be wrong.

In a typical fiscal year, the year over year change in average annual compensation for male Vice Presidents is quite different across Divisions. The annual change in compensation is about 45% higher, on average, in the Division with the highest growth (or smallest decline) in compensation than it is in the Division with the smallest growth (or largest decline) in compensation.<sup>16</sup> The largest difference of this sort occurred between 2008 and 2009 when compensation fell slightly for male Vice Presidents in Investment Management while average compensation increased by █████ in the Securities Division.<sup>17</sup> This is one more reason why all Divisions cannot be analyzed in a single model.

Likewise, controlling for whether an employee was hired as a lateral only accounts for the fact that, all else being equal, laterals earn more than non-laterals. The compensation of laterals, however, often involves guarantees and may be based on idiosyncratic market conditions and unusual productivity, such that the magnitudes of the factors that determine compensation are not the same for lateral employees as for internally promoted employees. In particular, "time at Goldman Sachs" cannot possibly have the same impact for a lateral Vice President hire as it does for someone promoted up through the Associate level.

Moreover, this is not simply a matter of opinion. There are standard tests that statisticians use to determine whether the models for two (or more) groups are sufficiently homogeneous that they can properly be combined in one statistical model, or, conversely, whether the models are so different that each group requires its own statistical model. These tests—known as Chow tests—are the routine, universally accepted method recognized in the econometric literature for testing whether different groups are properly or improperly combined in a single model. Performing this test is an indispensable step in ensuring the integrity of any statistical analysis that seeks to combine potentially disparate groups in a single regression.

---

<sup>16</sup> These numbers are based on the data in Dr. Farber's Initial Report; however, the numbers would be very similar regardless of the data used.

<sup>17</sup> 45% is the median difference between highest and lowest compensation growth among Divisions in the seven annual compensation changes in Dr. Farber's data.

The Office of Federal Contract Compliance Programs (“OFCCP”) in the US Department of Labor addressed this topic in its Notice on “Interpreting Nondiscrimination Requirements of Executive Order 11246 With Respect to Systemic Compensation Discrimination.”<sup>18</sup> On page 35139 the Notice recognizes that the factors that influence pay decisions may not necessarily have the same relationship to compensation for all categories of jobs. It identifies performance as one such factor and specifies “separate regressions for each category of jobs in which the relationship between the factors and compensation is similar” as one of the acceptable approaches to dealing with this problem. Another acceptable approach is to pool observations in the different categories and include “appropriate interaction terms,” which the OFCCP defines as a specification that allows the effect of performance (in this example) to have a different impact on compensation in each job category. The paragraph concludes “OFCCP will run statistical tests generally accepted in the statistics profession (e.g., the “Chow test”), to determine which interaction terms should be included in the pooled regression analysis.”

Dr. Farber could have conducted such “generally accepted” statistical tests to evaluate the validity of the many restrictive assumptions underlying his adoption of a simple statistical model applied to a heterogeneous set of employees. For example, he could have tested whether the effect of manager quartile on compensation was the same across all Divisions and the same for lateral and non-lateral hires. The tests to support or reject aggregation without interactions are a standard method in the econometric literature. For example, William Greene, in his textbook *Econometric Analysis*, Seventh Edition, describes these tests as a method to “reveal whether the assumed model had changed during the sample period, or was different for different groups of observations” (pg. 175). (As the passage from the OFCCP mentioned, they are often referred to as “Chow tests” by statisticians.)

Dr. Farber testified in his deposition (Farber Deposition 135-137) that he was aware that these statistical tests could have been performed to support his aggregation, but he did not conduct them. He gave no reason for his failure to do so. Applying these tests exposes a fundamental flaw in Dr. Farber’s work.

Under the Chow test, “p-values” above 0.05 indicate that the group(s) are similar enough to be combined in a single model. P-values below 0.05 indicate that the groups cannot be combined in a single model. For each of these tests that I perform on Dr. Farber’s regression models, the p-value that I calculate is *smaller than 0.0002*. In other words, the Chow tests unequivocally and resoundingly *reject* the proposition that these various groups can be combined into a single statistical model, as Dr. Farber does. Instead, the Chow results tell us that to produce a meaningful analysis, a

---

<sup>18</sup> Federal Register Vol. 71, No. 116, June 16, 2006; Notices: Page 35124 - 35141. Although this Notice was rescinded by the Obama Administration, it was in effect during the class period.



statistician must use (a) separate models for Associates and Vice Presidents;<sup>19</sup> (b) separate models for each of the Divisions; and (c) separate models for lateral hires and employees promoted from within the firm into their current level.

After this issue was brought to Dr. Farber's attention in my Initial Report, he partially, but insufficiently, heeds the Chow test results in his Rebuttal Report. In his Report, he now does separate studies for Associates and Vice Presidents, belatedly recognizing that these two types of employees cannot properly be grouped together. However, he inexplicably continues to combine all Divisions into a single model despite the Chow test's confirmation that this combination is statistically unsupportable.

The Chow test also tells me that, taken together, the effect on pay of all the factors in a statistical model is different for laterals than for non-laterals. Statistical tests require the separate analysis of these two groups. Dr. Farber does not do so.

That the relationship between compensation and the factors that determine it is not the same for each of these groups, is particularly important for the decision whether to combine employees in one large group or to analyze them separately. Conducting separate studies of the individual groups allows the contributions of relevant factors – production, experience, Business Unit, etc. – to vary across groups, as one would expect them to do. Failing to do this will distort the estimate of any gender difference.

In sum, Dr. Farber performed his analyses without using a routine and well accepted statistical test that would have told him that he cannot do what he does—combine very different populations for analysis in a single regression. His decision to separate just *one* set of these groups (Associates from Vice Presidents) in his Report, but not the others (Divisions and laterals/non-laterals), makes no sense, is inconsistent and statistically unsound.

## **V. DR. FARBER USES INCORRECT VALUES FOR COMPENSATION**

Dr. Farber testified that he intended to study annualized total compensation. He constructed annualized compensation using a field in the database labeled “total compensation for equity calculation,” which reflects compensation actually paid during the year. For employees who did not work a full year, he then adjusted this value for the fraction of the year that the employee did work. As Dr. Farber explained in his deposition (Page 80), “the idea was to look at the actual compensation paid to the individual in the year ... and then inflate that to what it would have been had they been there the entire year.” Dr. Farber's adjustment—which scales up all components of compensation by the same factor for partial year employees—distorts compensation for employees who were hired during the year and received

---

<sup>19</sup> In his Initial Report, Dr. Farber combined Associates and Vice Presidents into a single model. In his subsequent reports, he does separate models for Associates and Vice Presidents but does not separate out the different Divisions or lateral and non-lateral employees.



“sign-on” or “make-whole” bonuses (typically when they joined Goldman Sachs from a lateral position at a rival firm).<sup>20</sup>

This was wrong because make-whole bonuses are not pro-rated payments but are one-time-only payments.<sup>21</sup> A new lateral hire who received such a payment in the first half of the year would not receive another such payment in the second half (as Dr. Farber’s approach implies). Dr. Farber also testified that he believed that these bonuses were netted out of the year-end bonus. In reality, they have nothing to do with the PATC (Per Annum Total Compensation) or annual year-end compensation setting process. Consequently Dr. Farber’s adjusted annual compensation overstates PATC for employees who joined the firm mid-year and received a one-time bonus.<sup>22</sup> Because that group is more male than the population overall, this exaggerates any observed gender pay differences in Dr. Farber’s aggregate model.

## **VI. DR. FARBER’S STUDIES FAIL TO MODEL THE PROCESSES HE STUDIED**

In addition to refusing to apply standard statistical techniques to the work reflected in his Report, Dr. Farber’s regression analyses ignore factors that indisputably affect compensation and performance evaluations, and, instead, include factors for which there is no record support. As a result, the explanatory power of his studies is exceptionally weak.

### **A. Dr. Farber’s Use of Proxies for Productivity and Job Function in His 360 Review Process Regression Models Was Improper**

Dr. Farber’s 360 regression studies suffer from many flaws. For example, Dr. Farber controls for a series of factors that are typically found in academic research as *proxies* for productivity and job function; these proxies are used when the researcher does not have firm-specific data with which to create a model tailored to the specific employer’s processes.<sup>23</sup> See Tables 14 and 15 of the Report. Here, however, actual data on production and function were available; Dr. Farber ignores these data in his analyses, which is a fundamental flaw in his approach.

Dr. Farber does not identify any basis in the record for excluding the actual data on productivity and job function available to him and instead including proxy factors in his 360 review process regression. Dr. Farber uses Affirmative

---

<sup>20</sup> About 10% of the observations in Dr. Farber’s model are employees hired during the year who are potentially affected by this miscalculation.

<sup>21</sup> The “make-whole” bonus is a fixed amount paid to certain experienced hires who, for example, forfeit incentive compensation by leaving their previous employers. (Blinder Deposition, Pages 8 –11). In Dr. Farber’s statistics a sign-on bonus received at the half-year point is simply doubled under the assumption that it is pro-rated.

<sup>22</sup> Dr. Farber’s control for whether a person is a lateral hire in the current year (added belatedly to his Rebuttal Report studies) fails to correct this mistake.

<sup>23</sup> The factors that Dr. Farber uses in his 360 regression studies include: work experience at Goldman Sachs, work experience at Goldman Sachs squared, relevant experience prior to working at Goldman Sachs, relevant experience prior to working at Goldman Sachs squared, education, year, employee’s office and Division, whether a direct (i.e. lateral) hire, and Affirmative Action job group code.

Action Plan (“AAP”) job groups, for example, as a proxy for the function the employee performs. Yet Dr. Farber admitted at his deposition that he had no idea what these AAP job groups mean.<sup>24</sup>

Goldman Sachs does not use these classifications for any performance-related purpose, including the 360 review process. These AAP job groupings are broad generic descriptions of an employee's occupation, e.g., “Senior Finance Associate,” that cross Business Units and Divisions and are used, according to the regulations from which they are derived, for ascertaining the percentages of minorities and women who meet the basic qualifications to be deemed an applicant for particular types of jobs. These groupings do not reflect the different “roles” performed by Goldman Sachs employees after they are hired and when they are being assessed for performance, compensation or promotion purposes.

Dr. Farber's resort to the AAP job groupings remains unsupported by any evidence in the record of this case and was entirely unnecessary; he had job function information available to him in the data he was provided, but he chose not to use it. Dr. Farber's model is thus a fictional construct, with no basis or connection to the 360 review process.<sup>25</sup> He does not identify or explain how or why he chose these variables, and there is no sound basis for doing so.

Not only are the factors he uses unsupported, but Dr. Farber's reported results are less meaningful than he suggests. Dr. Farber reports differences that meet the technical definition of statistically significant but reflect differences that are quantitatively quite small. For example, in Table 15 of the Report, Dr. Farber reports differences between men and women in 360 review scores of .04, .05, and .06. These very small differences are made statistically significant, in large part, because of the large sample sizes. In other words, when the number of observations is very large, as it is when studying 360 scores, extremely small numerical differences will reach a level of statistical significance. Thus, the actual 360 review scores are only very slightly different on average as between men and women.

#### **B. Dr. Farber's Manager Quartile Regression Models Are Flawed**

Dr. Farber's manager quartile regression studies control for the same factors he uses in his 360 review regression studies, which means that these studies contain flaws similar to those discussed in the preceding section. See footnote 23 above and Tables 10 and 11 of the Report. The record in this case identifies the group within which the manager quartile is assigned: the Business Unit.<sup>26</sup> Yet Dr. Farber does not control for Business Unit in these regressions. The

---

<sup>24</sup> See Farber Deposition at pg. 110.

<sup>25</sup> When we look at the factors that Dr. Farber includes in his 360 model, it is not uncommon to see that they do not have the expected effects, or that they do not have statistically significant effects on 360 scores. Relevant prior experience is statistically significant in three of Dr. Farber's four 360 models, but it has a perverse effect. More relevant prior experience is estimated to *decrease* 360 scores. Education is only statistically significant with a pattern as expected in two of Dr. Farber's four 360 models. In one model education is insignificant, and in one its effects are not as expected (i.e., persons with Master's degrees are expected to have *lower* 360 scores than persons with Bachelor's degrees). This suggests that these factors do not belong in a proper model of the process at issue.

<sup>26</sup> See, e.g., Caroline Heller Deposition at pgs. 84, 88, 104, 320.

record also identifies important inputs to the manager quartile process: function, 360 review scores, and contribution or production where it is maintained and relevant to the employee's role.<sup>27</sup> These factors also are missing from Dr. Farber's work.

And there is no support in the record for the factors Dr. Farber does use, other than Division and year, as having any bearing on the manager quartile process. His use of the AAP job group is especially improper because these groupings include employees in multiple Divisions, and manager quartiles are not assigned across Divisions. Managers are encouraged to assign employees to quartiles roughly according to a predetermined distribution,<sup>28</sup> but employees in different Divisions do not compete with each other for spots at the top of the quartile distribution. Thus, an analysis of cross-Divisional groupings cannot capture the population that is compared for manager quartile purposes. Dr. Farber does not explain how he chooses his variables, why they would have any connection to the manager quartile process, or what statistical tests would suggest that his models include the proper variables. It is unclear that Dr. Farber even looked at the regression model results that underlie his opinion about gender differences since many of them make no economic sense at all.<sup>29</sup>

Further, as set forth in Part II, Section XII.C. below, my studies show that when I control for production and 360 review scores—the factors the record shows are the key inputs into the manage quartile process—there are *no statistically significant gender differences in any of the Divisions for either Associates or Vice Presidents*. See Tables 1 and 2 below. This finding remains un rebutted by Dr. Farber.

**C. When the Flaws in Dr. Farber's Revised Pay Regression Models Are Corrected, the Gender Differences He Reports Decrease Dramatically**

Dr. Farber's pay studies are similarly flawed. Dr. Farber does not include variables that are required by both statistical principles and by the actual processes used at Goldman Sachs to determine compensation. The data show that what an employee does and the concrete results he or she produces determine pay at Goldman Sachs to a large degree.

Dr. Farber ignores the available data reflecting these factors:

- Business Unit, where compensation budgets constrain actual pay decisions,
- Production, a measure of the actual financial contribution of the employee, and

---

<sup>27</sup> See, e.g., GS0109353-GS0109364, GS0109406-GS0109410; Caroline Heller Deposition at pgs. 344-45.

<sup>28</sup> See, e.g., GS0109353-GS0109364, GS0109406-GS0109410.

<sup>29</sup> Looking at the manager quartile assignment models that Dr. Farber estimates, it is immediately apparent that many of the factors included do not belong. In particular, relevant prior experience has a perverse effect (i.e., more relevant prior experience leads to worse manager quartile assignment) in all of the models. Education is insignificant in the model for Vice Presidents. For Associates it is significant, but more education generally leads to lower manager quartile assignment (i.e., Bachelor's degrees receive the best assignments, followed by Master's, followed by Ph.D., and followed by persons with multiple degrees).

- Function, which distinguishes groups of jobs actually performed at the firm.

First, Dr. Farber's statistical models ignore the fact that compensation decisions at Goldman Sachs are made at the Business Unit level. Each year, after each of the Divisions at issue receives its compensation budget from the Firm (based on the financial performance of the Division as a whole), Divisional management allocates money from that budget to its Business Units based on their contribution to Division and firm-wide success. In other words, in any year, the pool of money one Business Unit receives to pay its people can be vastly different from the allocation to another Business Unit (in absolute and per capita terms).

Employees in one Business Unit cannot be compared to those in another Business Unit where a different pool of money is being distributed, no matter how similar they may appear from other perspectives. As I testified at my deposition, when studying a personnel decision, factors that define the group within which the decision is made must be included in the model in order to compare similarly situated employees.<sup>30</sup>

Dr. Farber's model ignores these facts. His model anticipates that a man and a woman who are otherwise similar — but one of whom works in a Business Unit that has had a phenomenally successful year while the other in a Business Unit that struggled — will be paid the same, even though they are paid from different bonus pools of often dramatically different sizes. He then ascribes any difference he sees to gender rather than to the business realities underlying Goldman Sachs' approach to compensation.

Dr. Farber was asked at his deposition about the omission of Business Unit from his model. He acknowledged being aware that Business Units had budgets and that they made compensation recommendations based on those budgets (Farber Deposition, Page 130), and when asked whether the performance of individual Business Units would affect how people are paid, he replied "People may be paid quite differently for lots of reasons; that could be one of them." (Page 131) Nevertheless, he did not include Business Unit in his analysis.

Second, Dr. Farber fails to include important determinants of compensation in his model such as "production" — an individual employee's measurable contribution to the Firm's revenue. Dr. Farber states in his report that he intends to make pay comparisons among employees with similar "productivity-related" characteristics, but instead of using the direct measures of productivity available in the database, he instead relied on proxies for productivity — very weak and unreliable ones such as educational attainment and years of service at Goldman Sachs. At his deposition he

---

<sup>30</sup> Because pay decisions are made within the Business Unit and because compensation budgets are set at this level of Goldman Sachs, the most accurate way to estimate statistical models of pay would involve separate regression models for each Business Unit. However, I am not a proponent of following this strict methodology in a matter such as this as some of these Business Units are small, and when the number of observations is too small, a regression model may not be mathematically feasible; alternatively, the model may be unduly affected by a few observations.

acknowledged that he had the actual production data available to him, but he decided he would only use data fields that were available for every individual in his compensation analysis, testifying that he was only interested in production measures applicable to the entire class: “for m[e]—the population as a whole that I’m interested in, I can’t just have a productivity measure for some subset of them.” (Farber Deposition, Page 122). This is akin to refusing to consider productivity for factory floor workers because no similar production data were available for employees who design or market the product. As a matter of pure statistics, Dr. Farber’s approach does not make sense.

Whether production data is maintained for different groups of employees is, in many cases, a reflection of the different types of jobs performed. While production is not a meaningful concept for every job category, or every role in specific job categories, it certainly is for many, and it can provide important insight into performance for those employees—such as those in sales or trading jobs.<sup>31</sup> It makes no sense from a statistical perspective for Dr. Farber to have used “proxy” factors for production, such as education and years of service, instead of actual and available firm-specific production data.

Production is a key factor in explaining pay, especially in Securities, where a large number of the putative class members perform sales and trading functions, functions that allow for individualized attributed production.<sup>32</sup> My studies control for this factor. For those employees for whom this variable is available, production is highly important statistically in explaining compensation. Third, Dr. Farber’s model fails to account for an individual’s role or function within the Business Unit, even though roles reflect different skill sets, different external market values, and different relationships between compensation and its determinants.

Instead he used broad job groupings maintained for use only in government-mandated AAPs. These broad job groupings do nothing to refine the analysis. For example, the AAP job group called “Senior Finance Associate” includes employees in over 100 different Business Units and from every Division at issue in this case. Dr. Farber admitted that he did not make any reasoned, statistically-based decision that the AAP job group adequately modeled the work performed at Goldman Sachs; instead he acknowledged that he “never managed to find a satisfactory description of what the job groups actually were.” He testified that “[t]hey were numeric codes, but I don’t—I couldn’t tell you what any code was.” (Farber Deposition, Page 110). In other words, he used groupings to represent the employee’s role without knowing their

---

<sup>31</sup> For example, see the deposition of Caroline Heller, Pages 76 –78, 140 –145, for a discussion of production in Securities. At page 79 Ms. Heller discusses groups that do not have production. See also the deposition of Bruce Larson, Pages 87 – 90, 129 –143, and Exhibit 138 for production metrics used in IBD.

<sup>32</sup> This is not to say that productivity is irrelevant to explaining pay in other Divisions, only that the production data that the Securities Division maintains allows it to be more readily studied as a regression factor in statistical analyses for that Division.

definition, composition, or relationship (if any) to compensation decision making. Had he inquired, he would have learned that the AAP job group has no relationship to compensation decision making at Goldman Sachs.

It makes no sense for him to use AAP job groups as a proxy for job function in his pay studies, when those groupings are not meaningful collections of positions for purposes of compensation decision-making and when better data were available.

Dr. Farber's lack of concern about whether he incorporates all available information into his study is contrary to common statistical practice. According to the *Reference Manual on Scientific Evidence*, "Failure to include a major explanatory variable that is correlated with the variable of interest in a regression model may cause an included variable to be credited with an effect that actually is caused by the excluded variable. . . . In general, omitted variables that are correlated with the dependent variable reduce the probative value of the regression analysis."<sup>33</sup> Given that the deposition testimony and compensation documents identify 1) the Business Unit as the point of decision-making on compensation and 2) production metrics as a key consideration for many employees, Dr. Farber's decision not to consider these in his study renders his conclusions and inferences about gender differences unreliable.

Dr. Farber rejects the relevant and accurate data he had available to him on Business Unit, production, and function because he says they *may be* "tainted" and hence ought not to be included in his statistical measure of pay differences. See Farber Rebuttal Report at 42. According to Dr. Farber, any factor "may be tainted" if it is under the control of the employer. See Farber Rebuttal Report at 41.

Yet Dr. Farber has conducted no analysis to determine whether any of the factors that Goldman Sachs uses to manage its compensation processes is "controlled" by the firm or suspicious in any other way.<sup>34</sup>

Farber Deposition, Pg 92

Q. And so you don't have any opinion

14 that there's any taint or adverse effect from

15 that set of personnel decisions against women

16 in the group that you were studying, do you?

17 MR. KLEIN: Objection.

18 Q. You can answer.

19 A. That's correct.

Pg. 177-8

Q. Okay. You didn't do any taint

22 analysis of any of the processes or criteria

---

<sup>33</sup> Daniel L. Rubinfeld, "Reference Guide on Multiple Regression," *Reference Manual on Scientific Evidence: Third Edition* 2011, Page 314.

<sup>34</sup> Under this reasoning, the AAP job groups that Dr. Farber includes in his 360, manager quartile, and pay regression models could also be "tainted," as they are groupings that Dr. Farber alleges are "Goldman's own classification of affirmative action job groups." See Report at 20. Yet Dr. Farber continues to use this variable in all of his studies, even though it has no bearing on any of the processes he is studying. His insistence on using the AAP job group variable while he refuses to study the actual factors that influence pay and performance measures at Goldman Sachs is internally inconsistent.

23 that are used at Goldman Sachs to set  
 24 compensation or make promotion decisions, did  
 25 you?

3 A. I don't know what you mean by a  
 4 taint analysis.

5 Q. Well, any analysis that would lead  
 6 you to eliminate a factor in your equation  
 7 based on the fact that it was tainted, as  
 8 you've used the word?

9 A. It's a very hard thing to do, but I  
 10 did not do any of those analyses.

Dr. Farber is correct that it is a hard thing to do. But the first step in such an analysis would be to determine if the suspect factor is simply a disguised method for separating the outcome of men from that of women. When I look at the impact of these factors on pay for *just the men alone in the studied population*, I find that these factors are highly relevant to compensation, under circumstances in which gender cannot possibly have played any role. Thus, these critical factors (Business Unit, production and performance) have a statistically important impact in determining pay that has nothing to do with gender. Ignoring them, as he does, means that Dr. Farber's pay models bear no resemblance to the compensation processes at the firm.

When the fundamental flaws in Dr. Farber's models are corrected and the proper variables are included in the models, the gender differences Dr. Farber reports drop dramatically. For example, Dr. Farber reports an aggregate gender percentage difference in pay for female Vice Presidents of -21.8%. (Table 17, Model 2a of the Report.) When Dr. Farber's model is corrected—by separating out employees in different Divisions and hire status (lateral and non-lateral hires) as the Chow test requires, and adding in the variables that identify the employee's function, Business Unit, and production—

These three groups account for 74.4% of all women Vice Presidents studied in Dr. Farber's statistical models and have no significant gender differences in pay—even before any performance measures have been included. Thus, by correcting Dr. Farber's models but before accounting for either manager quartile or 360 reviews, the majority of women studied are paid no differently, statistically, than their peers.

For Associates, Dr. Farber reports an average pay difference of -7.7%. (Table 16, Model 2a of the Report.) When Dr. Farber's model is corrected as described above,

These groups account for 80.1% of all women Associates studied in Dr. Farber's statistical models and have no significant gender differences in pay—even before any performance measures have been included.



After adding in performance measures, *84.8% of the women analyzed in the regressions* are in a group for which gender differences are statistically insignificant.<sup>35</sup>

In sum, Dr. Farber continues to adhere to statistical models that have little or no relationship to the compensation process at Goldman Sachs. He does not incorporate the factors that are indisputably relevant to compensation at the firm, such as production and Business Unit, and he incorporates factors that have no connection to the processes he is attempting to model. When these flaws are corrected and the variables that actually determine pay are included in the models, there are no statistically adverse pay differences for the overwhelming majority of putative class members studied.

**D. Dr. Farber's Models Do a Very Poor Job of Explaining Pay Differences Within the Groups He Studies**

Correcting for the flaws in Dr. Farber's models not only decreases the reported gender differences dramatically; it also improves the explanatory power of the model itself. The "R-squared" is a measure of the explanatory power of a statistical model. This measure is an important piece of information because it shows how well a statistical model is able to explain variation in pay. Generally, the higher an "R-squared," the more variation in pay the model is able to explain. Dr. Farber does not report the "R-squared" for the models in his report. However, the backup data for Dr. Farber's regression model for Associates show an R-squared of 0.33, meaning that only 33% of the total variation in Associate pay, among men and women, is explained by the variables in his model. For Vice Presidents, the data show that his regression model has an R-squared of only 0.23.<sup>36</sup> These are low or "poor" R-squared measures, which indicate that his models are not doing a very good job of explaining variation in pay.

Once I correct the flaws in Dr. Farber's studies and add the factors that influence pay and performance at Goldman Sachs, the R-squared of the models improves dramatically. As shown in my tables discussed below, the R-squareds in my most complete models are all above 0.70, meaning that my models explain more than 70% of the variation in pay, which is *more than twice* the variation explained by Dr. Farber. My models "fit" the processes being studied far better than Dr. Farber's.

---

<sup>35</sup> These regression results are discussed in detail below.

<sup>36</sup> These R-squared measures are for Dr. Farber's preferred models, which do not include any performance measures. When he adds controls for 360 scores and manager quartile, a necessary step in building a proper model of the compensation process, the R-squareds of his models improve to 44% for Vice Presidents and 49% for Associates.



**VII. DR. FARBER'S GENDER DIFFERENCE IS NOT COMPELLING GIVEN HOW MUCH VARIATION IN PAY REMAINS UNEXPLAINED**

As I discuss in Section XIII below, compensation at Goldman Sachs varies dramatically, both from one person to another and for the same person from one year to the next. Table 3 in Section XIII illustrates the variation across Associates and Vice Presidents by comparing compensation at the 10<sup>th</sup> and 90<sup>th</sup> percentiles for the entire nation and Goldman Sachs. Among men in the US economy, compensation at these two points differs by about \$80,000, with the 90<sup>th</sup> percentile being 5.3 times the 10<sup>th</sup>. At Goldman Sachs, the difference is about [REDACTED] and the ratio is 8.9 to 1.

Most of the difference in pay between employees at Goldman Sachs consists of differences in discretionary bonuses determined at the end of the fiscal year. Developing a statistical study that captures a large percentage of this variation is challenging. Consequently, Dr. Farber's deficient study design explains little of the variation in pay of Associates and Vice Presidents. I use men-only studies to illustrate how little of the pay differences his regression models explain. Because the models are based on men-only none of the remaining pay differences can be attributed to gender. First, using a compensation model that includes only year, Division, office, AAP job group, and title (Vice President or Associate), I find that these factors account for 42.3% of the variation in pay among men.<sup>37</sup> Put another way, the bulk of the explanatory power in Dr. Farber's original model comes from the simple observations that: (i) Vice Presidents earn more than Associates,<sup>38</sup> (ii) New York employees at Goldman Sachs earn more than employees in other offices, (iii) some fiscal years were more profitable for Goldman Sachs than others, overall, and average compensation was higher in the more profitable years, and (iv) [REDACTED]

The controls to this point have only begun to bring the comparisons down to similarly situated men. People in the same title and Division can have different performance and different roles, and be in different Business Units. 57.7% of the original variation remains unexplained. Thus, I next use Dr. Farber's model to address variation in compensation *within* groups of men in the same Division, office, job group, and corporate title. I add education, tenure at Goldman Sachs, 360 score and manager quartile – the “productivity-related” factors in Dr. Farber's model. The addition of these factors raised the total explained variation in pay to only 58.9%.

---

<sup>37</sup> The percentages in this section are based on the studies in Dr. Farber's Initial Report, which combined Associates and Vice Presidents. However, the discussion is equally applicable to the updated studies in Dr. Farber's February 17, 2014 Report, which also leave a significant portion of variation in pay among men unexplained.

<sup>38</sup> After his Initial Report, Dr. Farber did separate regression studies for Vice Presidents and Associates.

This shows that Dr. Farber's purported "productivity-related" characteristics only explain 28.8% of the variation in men's pay that remained after the first step.<sup>39</sup> The remaining 71.2% of variation in compensation among men in the same peer group (defined by Division, office, AAP job group, and corporate title) is not explained by any of the factors that Dr. Farber has included in his model.

Consequently, Dr. Farber's ability to explain variation in individual pay with personal characteristics, as opposed to generic classifications, is quite limited. The vast majority of pay differences among Vice Presidents (or Associates) within the same Division and year is not explained by his model. In this context, Dr. Farber's estimate of the gender difference in compensation does not provide compelling evidence of gender discrimination.

The assumption in the unwarranted chain between Dr. Farber's estimate and an inference of gender discrimination is that none of the factors responsible for the enormous variation in compensation that he has been unable to explain are in any way different for men and women. This is an extraordinary leap. Dr. Farber's progression of studies shows that adding variables that explain variation in pay reduces the estimated gender difference. Given that pattern, and the large amount of variation remaining, one cannot easily dismiss the likelihood that the factors responsible for the yet unexplained pay differences among men will be correlated with gender and thus explain the remaining estimated disparity. Indeed, I show below that almost all of the remaining gender pay difference estimated by Dr. Farber is explained by factors that he omitted from his models.

#### **VIII. DR. FARBER'S MODEL PREDICTS SIMILAR PAY FOR EMPLOYEES WHOSE ACTUAL PAY IS VERY DIFFERENT**

A second way to assess how well the model fits the data is to compare the actual compensation of pairs of men that the model predicts should have similar compensation, given their characteristics. Again replicating Dr. Farber's most complete model, I identified 13,288 pairs of males with small (less than 0.5%) differences in predicted pay in each fiscal year. In other words, these are pairs of men whom Dr. Farber would predict to have identical compensation. Instead, I found that their actual compensation was very different: on average, the total compensation of the person with the higher pay in each pair of men was *2.11 times* the total compensation of the person with the lower pay.<sup>40</sup>

---

<sup>39</sup> Dr. Farber's so-called "productivity-related" factors account for 16.6 percentage points of a possible 57.7 percentage points of compensation variation within groups of men. Thus his model accounts for 28.8% (16.6/57.7) of within-group variation in compensation.

<sup>40</sup> The numbers in this section are based on the model in Dr. Farber's Initial Report. However, the substantive results and conclusions do not change when Dr. Farber's revised model is used.

Figure 1 shows the range of men's pay differences left unexplained by Dr. Farber's model.<sup>41</sup> The dots are individual compensation points minus the value of annual compensation predicted by Dr. Farber's regression.<sup>42</sup> This difference is the "error" or "residual" component of Dr. Farber's model. Points to the right of zero represent observations where an employee earned more than the model predicted, while points to the left of zero are observations where an employee's compensation was below its predicted value. Employee-year observations in the highest manager quartile are plotted at the top of Figure 1, followed by observations in lower manager quartiles in each successive row.

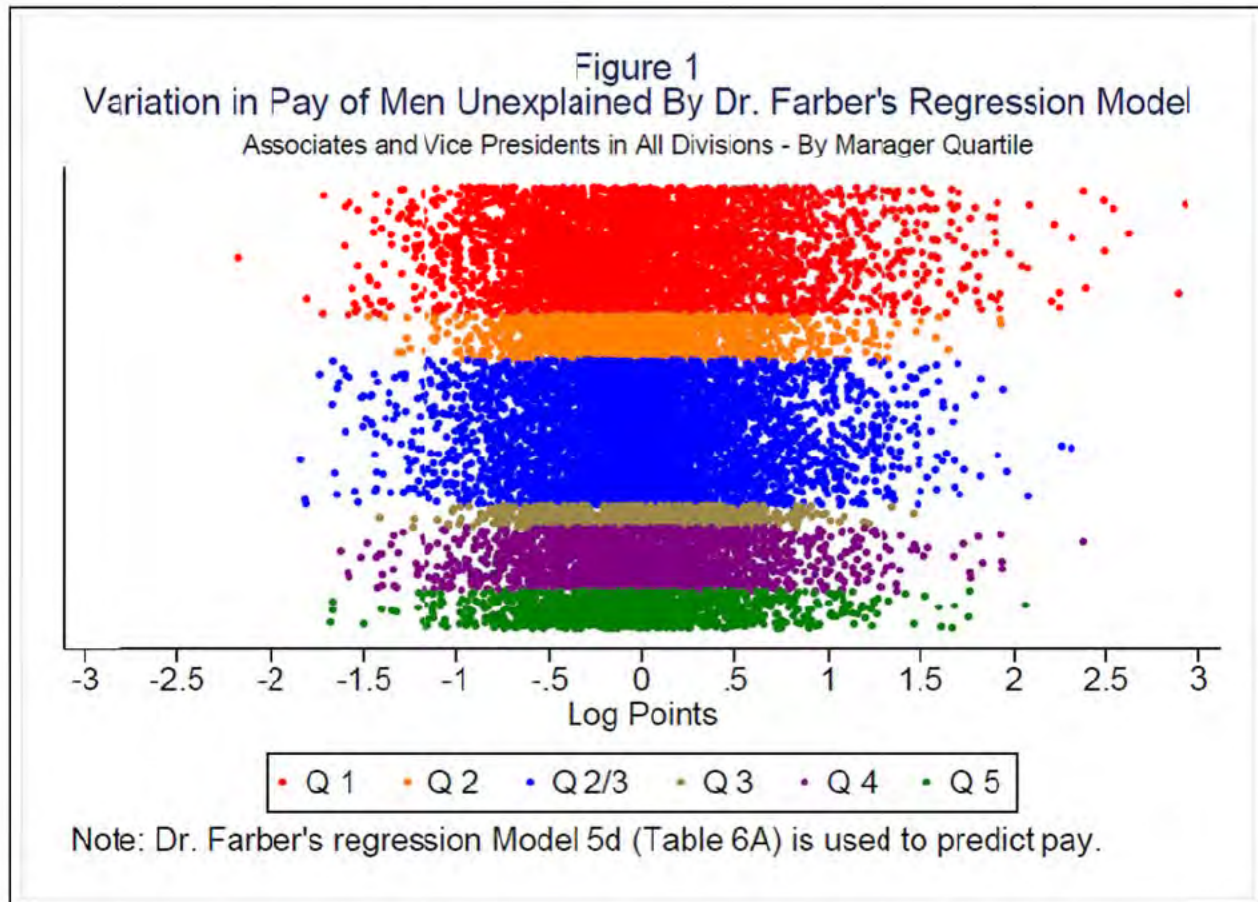
Figure 1 illustrates that the "errors" or "residuals" in Dr. Farber's model have a very wide range, in every manager quartile. Consider the compensation of an employee in the top manager quartile with residual earnings at the 90<sup>th</sup> percentile of the error distribution. This employee earns almost 4.5 times as much as the employee at the 10<sup>th</sup> percentile of the error distribution, after controlling for all of the factors in Dr. Farber's model. In the other manager quartiles, an employee at the 90<sup>th</sup> percentile of the error distribution earns about 3.1 to 4 times as much as an employee at the 10<sup>th</sup> percentile of the error distribution, again after controlling for the factors in Dr. Farber's model.

In other words, the wide variation in the error or residual component of Dr. Farber's model means that the predictive power of his model is very weak. His model does not explain why some men earn 3 to 4 times more than other men in the same manager quartile.

---

<sup>41</sup> This Figure was based on the model in Dr. Farber's Initial Report. However, the results and conclusions remain the same, even under Dr. Farber's revised model in his February 17, 2014 Report.

<sup>42</sup> Dr. Farber analyzes the logarithm of pay (as do I), so points here refer to log points.



#### Some Examples of Dr. Farber's "Unexplained" Pay Differences

I developed a sample of [REDACTED] pairs of Vice Presidents who, in 2011, had the same (or very similar) values for the factors used by Dr. Farber in his regression analysis.<sup>43</sup> I matched on the variables he controlled for, but I looked specifically for cases of people in different Business Units to try to understand how similar or different their roles and the factors affecting their compensation were. These are all people whom Dr. Farber would have considered "similarly situated."

<sup>43</sup> These matches were based on the models in Dr. Farber's Initial Report.

**IX. DR. FARBER'S CONCLUSIONS ABOUT THE IMPACT OF MANAGER QUARTILE AND 360 REVIEW PROCESSES ON GENDER DIFFERENCES IN PAY ARE OVERSTATED AND NOT WELL-FOUNDED**

There are two major failings about the manager quartile and 360 studies contained in Dr. Farber's Report: (1) they provide no evidence of common treatment of women in terms of 360 review and manager quartile processes across

the putative class; and (2) they should not be accepted because they are deeply flawed and unreliable according to well-accepted statistical principles. These fundamental observations are key to understanding the problems in Dr. Farber's conclusions about the impact of the 360 review and manager quartile processes on pay.

**A. The Impact of Manager Quartile and 360 Reviews on Pay Varies Widely Across the Putative Class**

Dr. Farber claims to show in his Report that the 360 review and manager quartile assignment results have a significant class-wide adverse impact on the pay of women. Dr. Farber, however, does not even study the issue of common impact. In fact, no such common impact exists. While manager quartile and the 360 review results have a highly statistically significant impact on the pay of all employees (and therefore these variables need to be included in a proper statistical model of compensation), the impact of these variables is not the same across Divisions, levels and lateral status.

I examine the effect of manager quartile assignment and 360 review scores on pay. Appendix Chart 6 displays the results of this study.<sup>48</sup> This chart shows that the impact of manager quartile on pay varies significantly based on the level of the employee and the Division in which he or she works. For example, for non-lateral Vice Presidents in Securities, assignment in the top manager quartile ("quartile 1") is estimated to increase an employee's pay by [REDACTED] relative to a similar employee in the next highest manager quartile ("quartile 2/3"). But for non-lateral Vice Presidents in IBD, assignment to the top manager quartile is estimated to increase pay by only [REDACTED] relative to a similar employee in the next highest manager quartile. In other words, being in the top quartile in Securities has [REDACTED] the impact on pay as being in the top quartile in IBD.

Similarly, the effect of 360 review scores on pay varies substantially across the putative class. For example, the impact of 360 review scores (prior to 2010) on pay varies from [REDACTED] among lateral IMD Associates to [REDACTED] among lateral Securities Vice Presidents. The effect of manager quartile assignment and 360 review results on pay is very different for employees in different Divisions and levels within the firm. There is no statistical evidence to substantiate the claim that these processes have a common impact on pay across the putative class.<sup>49</sup>

---

<sup>48</sup> Because the scale of 360 review scores changed in 2010, I estimate separate effects of 360 prior to 2010 and for 2010-2011.

<sup>49</sup> Appendix Chart 6 shows the wide variation of impact that the manager quartile and 360 processes have on pay over the entire population; however, I see the same variation in impact if I look at the pay of just women or just men.



**B. Dr. Farber Overstates the Importance of Manager Quartile Assignment and 360 Review Scores on Gender Differences in Pay**

As discussed above, manager quartile and 360 scores have a significant—but greatly varying—impact on pay generally. However, these factors have very little impact—certainly much less than reported by Dr. Farber—in explaining any reported *gender differences* in pay.<sup>50</sup>

Dr. Farber alleges that gender differences in the 360 review and manager quartile processes explain about 50% of the average pay differences for male and female Associates and approximately 22% of the average pay differences for male and female Vice Presidents. See Report at 31. These findings are very misleading. Properly constructed models—which correct for the flaws discussed above—actually show that little of the alleged gender differences in pay can be attributed to these two performance measures.

For example, when averaged across non-lateral Vice Presidents in all three Divisions, gender differences decline by an average of 2.4 percentage points when I control for manager quartile. For lateral Vice Presidents, the decline is 2.1 percentage points. This accounts for only a small portion of the average gender pay difference reported by Dr. Farber. When they are properly constructed, pay studies show a gender impact of manager quartile of less than two-thirds the size that Dr. Farber reports.<sup>51</sup>

The same is true for Associates. The reported gender differences for non-lateral Associates only decline by an average of 2.5 percentage points when I control for manager quartile. For lateral Associates, the decline is 2.0 percentage points. This is smaller than the impact on gender pay differences that Dr. Farber reports.

Thus, while manager quartile is very important in explaining pay generally, it is much less important in explaining the average pay differences between the genders. That is because, as the data shows, the average gender difference in manager quartile assignment is small.<sup>52</sup>

As is true of manager quartile, the impact of 360 review scores on gender differences in pay varies by Division, level, and hire status (lateral or non-lateral) and is quite small. In all three Divisions, the gender difference for non-lateral Vice Presidents only declines by an average of 1.3 percentage points when I control for 360 review scores. For lateral Vice Presidents, the average gender difference *increases* by 0.1 percentage points. Thus, gender differences in 360

---

<sup>50</sup> This is because other key factors (such as Business Unit, production, and function), which Dr. Farber ignores, have a much greater impact on any alleged gender differences and because the difference in manager quartile assignments between men and women are small.

<sup>51</sup> I look at manager quartile first, before 360 review scores, because it is a factor taken into account in the compensation process. 360 review scores are an input into manager quartile; I look at this measure separately because it offers a more granular form of feedback with its more precise scores.

<sup>52</sup> For Vice Presidents in Dr. Farber's models, 28.3% of men and 25.8% of women have the highest (top) manager quartile without any adjustments for any other factors.

reviews have a small incremental effect on gender differences in compensation when analyzed within the parameters of a proper compensation model.

Dr. Farber improperly attributes a high proportion of the gender pay differences he reports to manager quartile assignment and 360 review scores, but properly modeled studies refute this assertion. In sum, the most important factors in explaining gender differences in pay are those that Dr. Farber ignores: Business Unit, lateral status, function, and production. By contrast, only a small and varied percentage of the pay differences can be attributed to these performance measures.

**X. DR. FARBER'S PROMOTION STUDIES ARE INCONSISTENT WITH HIS OTHER STUDIES; WHEN PERFORMED CONSISTENTLY WITH HIS OTHER MODELS, THE RESULTS ARE NEUTRAL TO WOMEN**

In his promotion analysis, Dr. Farber adopts a different statistical methodology than he uses in his compensation analysis. He does not control for 360 score or manager quartile, and he uses a different statistical framework to test the statistical significance of the gender difference he purportedly found. Had he used the same methodology in his promotion analysis as he did in his study of compensation, he would have found no significant difference between men and women in promotions from Vice President to Extended Managing Director ("EMD").

First, Dr. Farber's promotion study is inconsistent with his own stated methodology for studying processes where the outcome is binary, that is, "promoted" or "not promoted." In paragraph 60 of his Report, where Dr. Farber explains how he analyzes differences in quartile assignment, he states: "The statistical model used here is called a probit model . . . [which] is appropriate for cases where the outcome variable takes on two discrete values (in this case, placement in the top quartile or not)."

Yet, Dr. Farber does not follow his own recommended model when studying promotions, which also involve a discrete or binary value—either promoted or not promoted. Normally one would expect a statistician to use the same methodology when the same type of decision-making process is being analyzed, as is the case with promotions and assignment to the top manager quartile. Dr. Farber does not do so. Instead, for his promotion study, Dr. Farber inexplicably looks only at the promotion rates for men and then calculates the expected number of promotions for women based on the male selection rate. This is not only inconsistent with his binary outcome model; it also assumes no limit on the number of promotions that can be awarded. This is not an accurate reflection of the promotion process because candidates compete for a limited number of slots, as is the case with assignment to the top manager quartile. If we assume 100 equally qualified candidates, half men and half women, and ten available promotions, we would expect that a neutral process would result in five promotions of each gender. However, if six men were promoted, Dr. Farber's model



would expect six women to be promoted. This approach is not consistent with a process in which there are limits on the number of possible promotions.

Second, in his compensation model 2d, Dr. Farber includes the adjusted average score from the 360 reviews and the manager quartile as additional measures of “productivity-related” characteristics. Inexplicably, however, he uses neither the 360 review data nor the manager quartile in estimating who is more likely to be promoted. Dr. Farber testified in his deposition that he did not give much thought to the factors he included in his promotion analysis.<sup>53</sup>

Third, Dr. Farber’s statistical framework for promotions is inconsistent with his compensation analysis. In his compensation study, Dr. Farber estimates a single model for both men and women and includes an indicator for the employee’s gender. His estimated gender pay differences were simply the “coefficient” on the gender indicator variable. (Farber, Paragraphs 38-41) In his promotion analysis, Dr. Farber estimates a model for men only and then predicts how many promotions would have occurred for women if their progression was identical to the progression he estimated for men. The single-step framework he uses for compensation and the two-step process he uses for promotions are similar, but their results when applied to promotions are different. His use of different models when nothing in the data or the questions presented would require it is inexplicable.

Using the two-step framework, Dr. Farber finds a shortfall in promotions of women that is statistically significant. Had he estimated a promotions model using the same variables and single-step framework that he used for compensation, the estimated shortfall in promotions for women would have been statistically insignificantly different from zero (-1.72 standard deviations). Moreover, had he used performance metrics and all the years of the data, he would have found that women are promoted more often than men.

A final concern with Dr. Farber’s promotion study is that he limits the study period to the period from 2004 to 2008. He stated in his deposition (Page 125) that this was requested by Plaintiffs’ Counsel and that he did not know why. Thus, Dr. Farber’s study leaves out the last three years of promotions in the putative class period. Again, this is inconsistent with his study of pay and his rationale for this different approach, if there was one, is unstated.

Dr. Farber should look at actual promotions for the entire population, as I do, rather than creating a false benchmark based on male promotions alone. Had he done so, he would see that the results are statistically neutral for

---

<sup>53</sup> “I don’t—you know as I sit here today, I don’t have a strong recollection except I kind of did this analysis, I want to say, independently of the pay analysis. I basically said, let’s do an analysis [of] promotion, what education you’re in. I thought Division would be important. And the time as VP as meant to proxy of how long you’ve been in your current position which is not a variable certainly I include in the pay analysis. So, there’s a variable here that’s not in the pay analysis. And now that I look at it, I see I don’t have time – overall time at Goldman Sachs and I don’t have job group.” (Page 143)

women in promotions. His own reported results would fall to below 2 standard deviations and he could not report the results shown in his Report.

In sum, there are numerous flaws in Dr. Farber's work. His choice of models, his refusal to use available explanatory data, and his presentation of a single firm-wide result for pay and promotion without examining whether such aggregation passes professional tests for its appropriateness, lead me to conclude that his studies are unreliable as a basis for any conclusions about alleged gender differences in personnel decision-making concerning the putative class members.

## **PART II – MY STUDIES AND ANALYSES**

As explained above, the work performed by Dr. Farber is seriously flawed because it fails to account for the way that compensation,<sup>54</sup> performance review, and promotion decisions are made at Goldman Sachs. The design of Dr. Farber's statistical models obscures rather than illuminates the key statistical questions: is there pay discrimination as alleged by Plaintiffs and are any observed differences common to the members of the putative class?

When a statistician or labor economist is asked to do a statistical analysis of compensation, the basic approach is the same: to build a model that, as nearly as possible, accounts for the way in which pay is actually determined in that organization. That means the investigator must understand and account for the pay process — the structures and mechanisms used to determine pay — and for the factors that normally would be expected to influence why people are paid what they are paid.

For example, if I am studying a manufacturer I would want to know whether some or all of the employees are paid based on the volume of production of the goods being manufactured. For those who are paid in this way, my model would need to account for productivity as one key variable. For those who are not — administrators or managers, for example — I would need to understand what behaviors or attributes are rewarded in their jobs, such as seniority or education, and the competitive environment for their talents. Dr. Farber's models do not account for the determinants of pay at Goldman Sachs.

I develop my models in a different way. First, I design compensation models that are consistent with the decision-making processes at Goldman Sachs. I show that these separate models must be used to reflect properly the differences between Associates and Vice Presidents, and between Divisions. I look at the inputs the record demonstrates are used for compensation decisions at the Firm, prime among them the 360-degree performance review scores ("360 scores"), the

---

<sup>54</sup> As I understand it, Plaintiffs do not allege discrimination with respect to their base salaries, and only the incentive compensation awards to employees in the three Revenue Divisions are at issue here.

“manager quartile”<sup>55</sup> assigned by the employee’s manager to measure absolute and relative performance of the employee, and the metrics reflecting revenue production for those employees for whom such data is recorded. My studies show that, contrary to Plaintiffs’ allegations, managers assign manager quartile scores to their subordinates on a gender neutral basis that is highly correlated with employees’ 360 scores and the revenue production numbers. Indeed, looking at Business Units (the level at which dollars are allocated for compensation purposes), and accounting for those factors, I find there are only small gender differences in pay across Divisions and Business Units, and, more importantly, that those differences sometimes favor men and sometimes favor women. There is no pattern of pay differences adverse to women.

**XI. THE STRUCTURE OF THE FIRM AND THE PROCESS BY WHICH COMPENSATION DECISIONS ARE MADE AT GOLDMAN SACHS MUST SHAPE ANY PROPER STATISTICAL ANALYSIS**

To understand the organization of the Firm and its Divisions, which is essential to construct a reliable statistical model, I reviewed deposition testimony from all of Goldman Sachs’ corporate witnesses, including those specifically designated to respond to questions regarding the organizational structure of each Division.<sup>56</sup> I also reviewed the depositions of all witnesses designated to discuss compensation practices at the Firm and in the Divisions.<sup>57</sup> I reviewed voluminous documentary evidence, including organizational charts, compensation guidelines and other materials detailing the structure of the Divisions and Business Units. Additionally, I reviewed portions of the transcripts from Plaintiffs’ depositions.

Goldman Sachs is divided into separate Divisions, including the three revenue-producing Divisions that are at issue: Securities, Investment Banking, and the Investment Management Division. Each Division has different areas of focus and expertise and provides different services. Each Division consists of numerous Business Units, totaling more than 100 across the three Divisions.

One of the principles of compensation at Goldman Sachs is that total compensation for an employee (other than those paid on straight commission) is determined by the Firm’s performance, the Division’s performance, the Business

---

<sup>55</sup> Although referred to as a quartile, manager quartile does not divide persons into equal fourths. The bottom quartile is always split into category 4 (the 25th to 10th percentiles) and category 5 (the bottom 10 percent). Furthermore, in many years and Divisions the middle 2 quartiles are combined into a “2/3” category, which includes all persons between the 75th and 25th percentiles. Landman Deposition, Pages 20-22; 52-53. In some years, assignment to categories 1 (the 1st to 25th percentiles) and 5 are the only assignments that are consistently recorded. In some years and Divisions assignments to “2/3”, and sometimes 4, are not recorded. In my analysis, I always combine 2s and 3s into the “2/3” category. In year-Divisions for which category “2/3” is not recorded, I assign a category of “2/3” to persons who are included in the population of people to be quartiled but who were not assigned a category of 1, 4, or 5.

<sup>56</sup> See depositions of Rodney Miller, Gargi Banerjee, Beth Cogan, Scott Mehling, and Genevieve Felix.

<sup>57</sup> See depositions of Caroline Heller Sberlati, Bruce Larson, Jessica Kung and Lisa Donovan.

Unit's performance (including the employee's department or desk), and the individual's performance.<sup>58</sup> Annual compensation budgets are determined near the end of the fiscal year, when the financial performance for the Firm, Divisions, and Business Units is reasonably well known. (Mehling Deposition, Page 64-65). Because the final results for the fiscal year are not available when the compensation process begins each year, there are often multiple compensation "rounds." Compensation of individual Associates and Vice Presidents is proposed in the first round and then, in subsequent rounds, revised within the Divisions. My discussion here focuses on the process in the first round. The testimony indicates that processes in any subsequent rounds are consistent with those of the first round.<sup>59</sup>

The first step in the year-end compensation process occurs when the Firm's senior management determines what total Firm-wide compensation will be and how the Firm-wide total will be allocated among the Divisions. (Mehling Deposition, Page 64-65). The Division heads then determine how much money will be available to each Business Unit within their respective Divisions. The Business Units, of course, must pay their employees within the budgets they are provided.<sup>60</sup>

Once the initial allocations to the Business Units have been made, Business Unit leaders further allocate budgets to the various departments that make up the Business Units. After receiving their budgets, managers within the Business Units make compensation recommendations for employees who report to them. Those recommendations are based in part on the manager quartile placement of the employees, a decision which is also made at the Business Unit level.<sup>61</sup> The manager quartile, which is an assessment of employee contribution and potential, relies heavily on the results of the 360 review process and on production numbers, where present, as well as a host of other factors.<sup>62</sup> Along with the quartile, compensation recommendations are based on market and other factors.

The managers' compensation recommendations are discussed with higher level managers and Business Unit leaders, all within the Business Unit. Finally, each Division also has its own compensation committee, which is made up of Business Unit leaders and other senior Divisional management.<sup>63</sup>

---

<sup>58</sup> See, for example, the first page of Heller Deposition Exhibit 141, "2011 Investment Banking Division Compensation Guidelines: Compensation Managers." (GS0113426) This language is used in the 2011 Compensation Guidelines for the other three Divisions. Similar language is found in the Compensation Communication Guides for the various years. In 2002, the guide begins "It is important to note that the total compensation for each individual employee reflects a combination of performance of the firm and Division for the fiscal year, and the employee's performance in the context of their role." (GS0113712)

<sup>59</sup> Heller Deposition, Pages 28-33; Larson Deposition, Pages 65-66; Kung Deposition, Pages 112-115.

<sup>60</sup> Heller Deposition, Page 179; Larson Deposition, Pages 61; Kung Deposition, Pages 57-59.

<sup>61</sup> Heller Deposition Exhibit 191; Larson Deposition Exhibit 141; Kung Deposition Exhibit 220; Donovan Deposition Exhibit 170; GS109388-9; GS0115071-2.

<sup>62</sup> Donovan Deposition Exhibit 132.

<sup>63</sup> Committee members are provided with information regarding all individuals in the Division, but the focus of their meetings is on the performance of the individual Business Units and departments; during the meetings, only top earners in the Division are typically

Because manager quartile decisions and, then, compensation recommendations are made within the Business Units, and because the Business Units in each Division operate under entirely separate compensation budgets, the Business Units must be analyzed separately. As discussed, Dr. Farber, who does not even perform analyses by Division (and instead merely controls for Division), clearly disagrees. However, I believe his disagreement is based in part on his lack of knowledge of Goldman Sachs' structure and processes.<sup>64</sup>

## **XII. PLAINTIFFS' CRITICISMS OF THE FIRM'S MANAGER QUARTILE ASSESSMENTS ARE NOT SUPPORTED BY THE DATA**

Plaintiffs claim that the manager quartile process is a "common" practice that results in supposed compensation and promotion disparities between similarly situated men and women. In this section I show that (i) manager quartile assignments are strongly related to job performance metrics, (ii) the manager quartile process varies among Divisions, and (iii) the process is gender neutral when comparing quartile placement of similar employees.

First, I consider whether manager quartile assignments are disconnected (i.e. "largely untethered") from 360 performance review scores, as Plaintiffs allege, and whether manager quartile assignments are related to "production oriented" performance measures. Next I explore whether the manager quartile process generates statistically significant outcomes that are adverse to women once these factors are taken into account. Finally, I provide evidence that the manager quartile process varies across Divisions.

### **A. Manager Quartile Assignments are Significantly Positively Related to 360 Reviews**

Plaintiffs allege that the manager quartile review process is disconnected from the results of the 360 review. My analysis of the data refutes this notion and instead shows that manager quartile assignments are significantly related to 360 scores.<sup>65</sup> My empirical results demonstrate that, far from being disconnected, the 360 review process and the manager quartile process are closely related to one another.<sup>66</sup>

---

discussed. While there are often changes in individual compensation amounts between the different "rounds," of compensation, the evidence demonstrates that the most common reason for a change is that either more or less money is being allocated to the business; in those situations, the compensation committees do not mandate individual changes. The lower level managers within the Division and Business Units make the revised recommendations in light of the budget changes. Heller Deposition, Pages 23-25 and 28-33; Larson Deposition, Pages 62-63 and 65-66; Kung Deposition, Pages 64-66 and 112-115.

<sup>64</sup> When he was asked his opinion about possible variability in differences by gender by Business Unit, he responded:

A: "I am not sure I understand what you mean by business unit."

Q: Business unit is an organizational level at which the budgets are allocated for compensation setting. So –

A: Can you give me an example?

Q: Sure. America's One Sales.

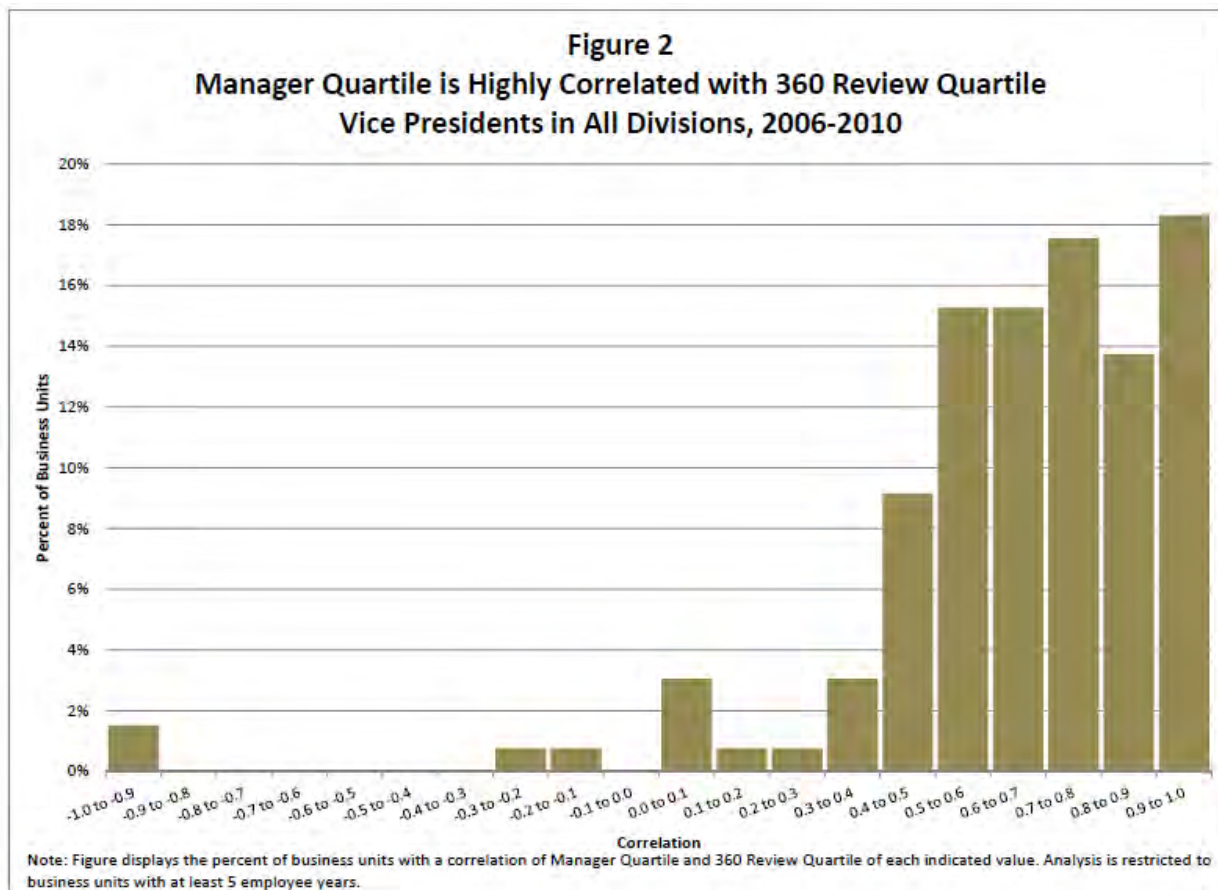
A: That doesn't help."

Farber Deposition, Page 228.

<sup>65</sup> I calculate correlations among employees who are both included in the FRS (360 Review) data and receive a manager quartile assignment in the same year between 2006 and 2010. I do not include 2011 in the analysis because the calculation of 360 review quartile in 2011 differs from previous years.

<sup>66</sup> Dr. Farber did not analyze the relationship of manager quartile to performance-related metrics.

If manager quartile assignments were “untethered” from 360 reviews as Plaintiffs allege, I would expect to observe an insignificant association between manager quartile assignments and 360 review scores. For example, if the processes were disconnected or “untethered,” knowledge that an employee review score was in the top 360 review quartile would provide no valuable information about the employee’s manager quartile placement. That is, there should be no statistical pattern of an association between higher or lower manager quartile assignments and higher or lower 360 review scores or quartiles.



Instead, Figures 2 and 3 show a dramatic statistically significant positive correlation between 360 score and manager quartile assignments within a Business Unit for employees with the same corporate title<sup>67</sup> (e.g., for Vice Presidents in FICC Americas Sales). Correlation coefficients for Vice Presidents are presented in Figure 2 and for Associates in Figure 3. A correlation coefficients measure the association between 360 review and manager quartiles. A positive correlation coefficient means that a higher 360 review score is associated with assignment to a higher manager quartile, and a negative value indicates that a higher 360 review score is associated with a lower manager quartile

<sup>67</sup> In the Investment Banking Division I group employees by Junior Banker and Senior Banker rather than the corporate titles of Vice President and Associate.



assignment. A correlation of zero would indicate no significant relationship between the two performance measures.<sup>68</sup> The 20 categories (or “bins”) included in each chart are defined in units of .10 ranging from -1.00 to +1.00 (1.0 meaning a perfect, positive correlation, and -1.0 meaning a perfectly inverse correlation). The categories aligned to the right of each chart show the Business Units with the strongest positive correlations between 360 reviews and manager quartiles, while categories to the left indicate the Business Units with the strongest inverse relationship. These measures would be insignificantly correlated (i.e., the correlation coefficients would be close to zero) if manager quartile assignments are disconnected from 360 reviews, as Plaintiffs assert. However, Figure 2 indicates a significant positive correlation between 360 review quartile and manager quartile for Vice Presidents in almost all Business Units. In Figure 2, the pairwise comparisons of manager and 360 Review quartiles represent 131 Business Units, of which 127 comparisons result in a positive correlation and 106 of these positive correlations are statistically significant.

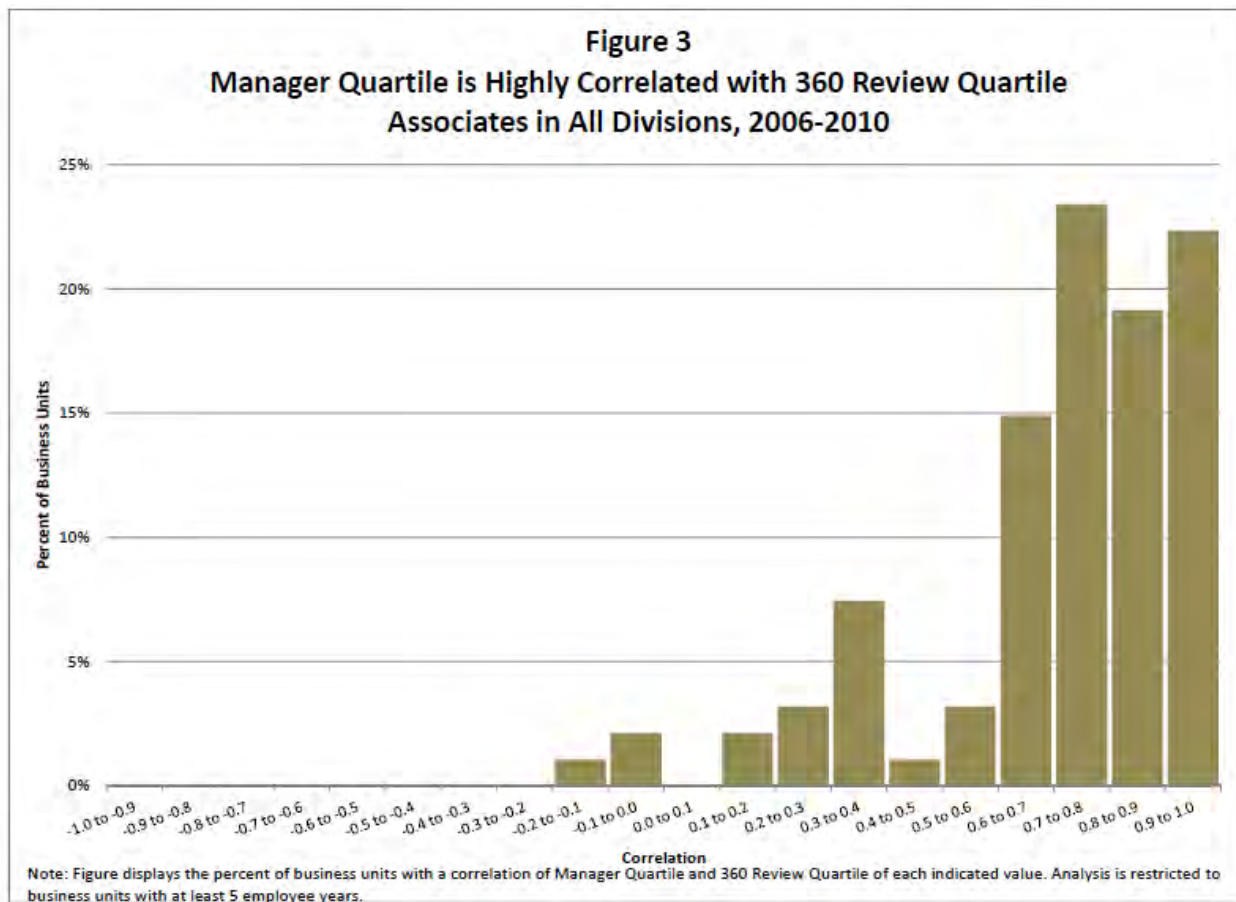


Figure 3 presents the distribution of correlation coefficients between manager quartile assignments and 360 reviews for Associates within the same Business Unit. I find that the correlation between manager quartile and 360

<sup>68</sup> In statistical analyses the maximum possible correlation coefficient is 1 (and the minimum possible value is -1) and a correlation coefficient of zero means the absence of a relationship between the variables.

review quartile is positive in 91 of 94 Business Units.<sup>69</sup> For both Vice Presidents and Associates there is an overwhelmingly significant positive correlation between the outcomes of the manager quartile and 360 review processes.

**B. Manager Quartile Assignments are Significantly Related to “Production Oriented” Measures of Performance As Well**

I next examine the relationship between manager quartile assignment and an employee’s “production oriented” measure of performance in the Securities and Investment Banking Divisions (where I can observe individualized “productivity related” performance measures).<sup>70</sup> I construct production quartiles based on an employee’s production in each Division and fiscal year. I assigned employees with production in the top 25% of employees with the same corporate title in the same Division and fiscal year to the top production quartile, while employees whose production ranked in the next 25% were assigned to the second quartile, and so on. I calculate these correlations for employees with the same corporate title within each Business Unit among employees with “production oriented” measures of performance. If the manager quartile process was disconnected from production oriented measures of performance as Plaintiffs allege, I would expect a weak correlation between manager quartile and production quartile assignments.

On the contrary, Figures 4 and 5 present evidence of a positive relationship between an employee’s production quartile and manager quartile assignment within Business Units. The histograms in Figures 4 and 5 are analogous to the ones in Figures 2 and 3. A positive and significant correlation coefficient means that employees who rank higher in terms of “production oriented” performance within their Business Unit receive higher manager quartile assignments, on average. A negative correlation indicates that employees with lower production tend to be assigned to a higher manager quartile.

Figure 4 indicates that the correlation between production quartile and manager quartile is positive in all 50 Business Units for Vice Presidents, and it is statistically significant in 30 of these Business Units. Figure 5 shows that the correlation between manager quartile and production quartile is positive in 46 of 49 Business Units and statistically significant in 25 of these Business Units for Associates.<sup>71</sup> The evidence strongly contradicts Plaintiffs’ allegations that manager quartiles are disconnected from 360 reviews and “production oriented” measures of performance.

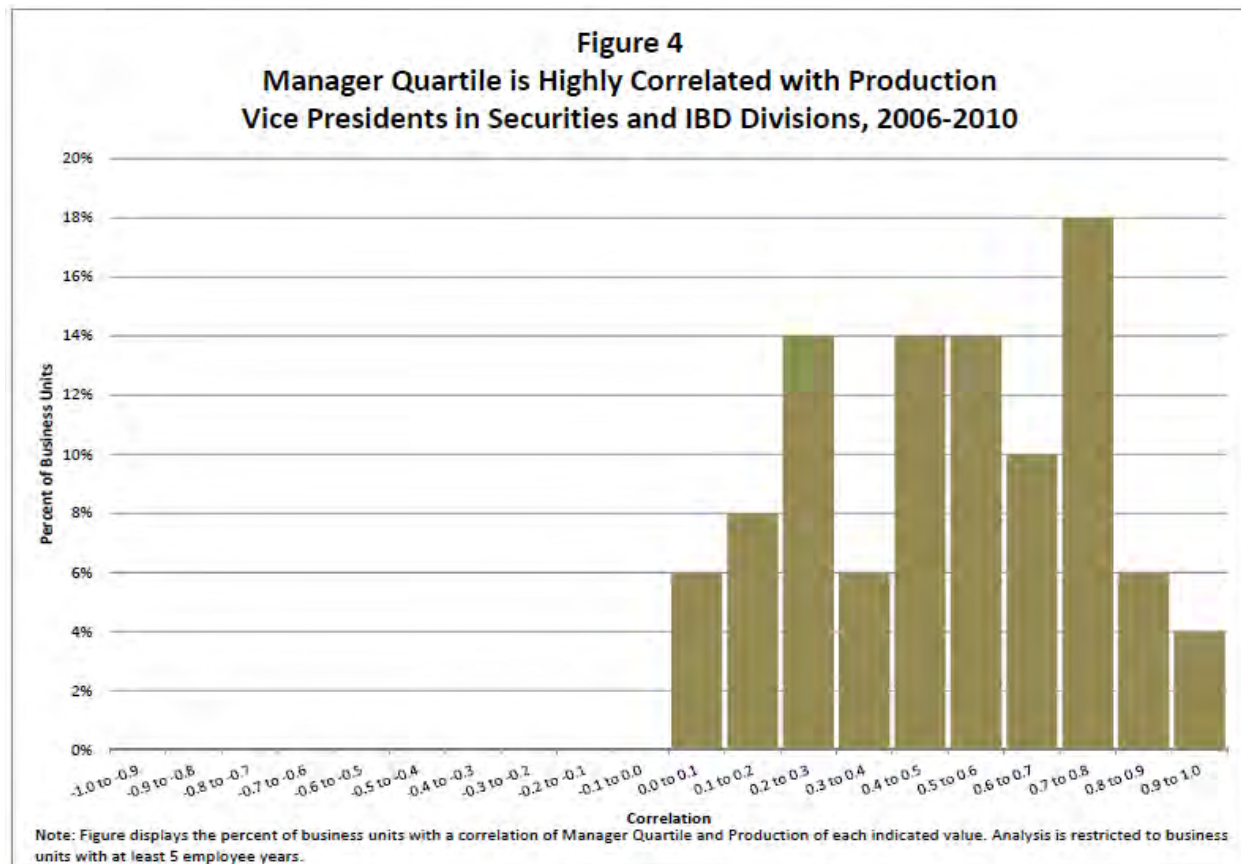
---

<sup>69</sup> I also examine correlations between 360 review scores and manager quartiles by Division. I find a significant positive association between 360 review quartile and manager quartile assignments in the vast majority of Business Units in all Divisions. The correlation tends to be slightly higher, on average, in Investment Banking and slightly lower, on average, in Investment Management, relative to the other Divisions.

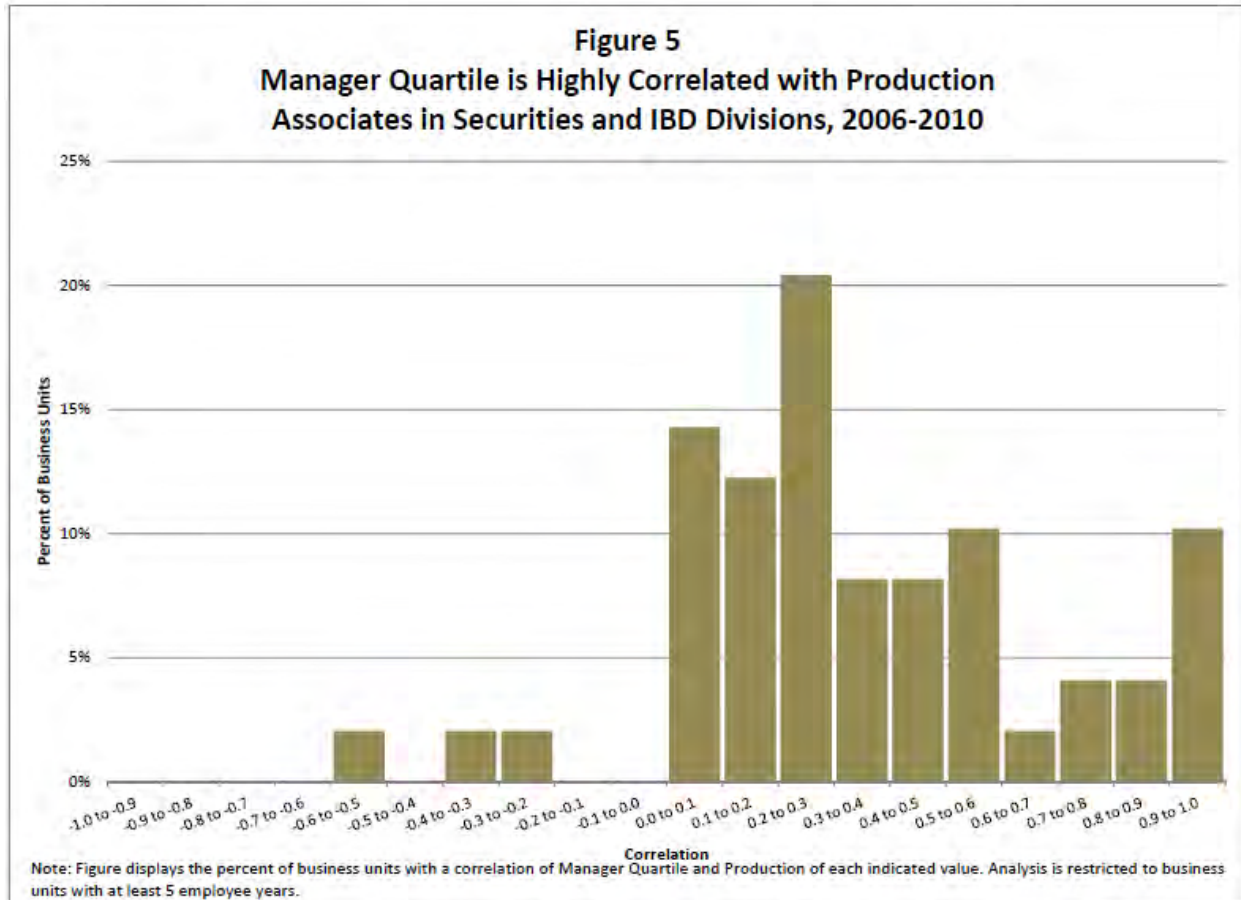
<sup>70</sup> Raw individualized production measures differ between Investment Banking and Securities. I grouped production into quartiles to facilitate comparisons across Divisions, corporate titles, and Business Units.

<sup>71</sup> I also examine patterns of correlations between manager quartile and production quartile by Division and found a clear pattern of significant positive correlations in both Securities and Investment Banking. The magnitude of the correlation between production and manager quartile differs between Divisions; the correlation is slightly smaller and less statistically significant, on average, in Investment Banking than in Securities.





While the correlation between both 360 reviews and “production oriented” measures of performance and manager quartile assignment are most often positive and significant, the magnitude of the correlation differs across Divisions. This finding is consistent with my understanding that managers in different Divisions put a different weight or emphasis on alternative performance measures when making manager quartile assignments. The process of assigning employees to manager quartiles is not formulaic. The lack of “commonality” is reflected, in part, by the variety of correlations between performance measures and manager quartile assignments presented in Figures 2-5. While there is a clear positive association between the manager quartile process, 360 reviews, and “production oriented” measures, managers in different Divisions weigh alternative performance metrics differently when making manager quartile assignments.



### C. The Manager Quartile Process is Gender Neutral

Plaintiffs allege that the manager quartile process “is a common, unvalidated procedure through which discrimination occurs.” For this allegation to be true, the manager quartile process would operate in a systematic manner and result in consistent and statistically significant adverse outcomes for women. Hundreds of managers in the Business Units participate in the manager quartile process every year, and thousands of different managers participated in the process across the proposed class period.

If the manager quartile process was a means to discriminate against women I would expect to see a systematic pattern where men are statistically significantly more likely to be assigned to the highest manager quartile and statistically significantly less likely to be assigned to the lowest manager quartiles, relative to comparable women. That is not the case.

I use two types of manager quartile assignment models; one model explains assignment to the highest quartile and another explains assignment to the two lowest quartiles.<sup>72</sup> I use these models to test the hypothesis of gender neutrality in the manager quartile process after controlling for individual differences in performance.<sup>73</sup>

I estimate a total of 12 assignment models, four models for the Securities Division, four more for IMD and four for IBD. Within each Division, I estimate separate models for senior Vice Presidents and Associates (Senior Bankers and Junior Bankers in IBD). Separate models are appropriate because statistical tests reveal that the manager quartile process differs significantly by Division. Individual managers and decision-makers in different Divisions put different weight on the available performance metrics when assigning employees to quartiles.

I use the same explanatory factors for assignment to the highest and lowest manager quartiles as described above: an employee's 360 review score (including adjustments for changes in the 360 review score range by fiscal year), production quartile, Business Unit, fiscal year, and function (in the Securities Division).

The models described above provide twelve different estimates of possible gender differences in manager quartile assignments. None is statistically significantly different from zero. All of the results are consistent with gender neutrality in decision-making.

Table 1 presents gender differences in assignment rates to quartile 1 based on the statistical model. For example, for Associates in Securities the model indicates that women are 0.19% more likely to be assigned to the top manager quartile than men in the same Business Unit with the same production and 360 review quartile. This gender difference is insignificantly different from zero and equivalent to a difference of 0.08 standard deviations. The remainder of Table 1 presents the corresponding estimates for the other corporate titles and Divisions. While the models I estimate show a significant relationship between some factors, including 360 reviews, and manager quartile, gender differences in assignment to the highest manager quartile are insignificantly different from zero in all Divisions for all corporate titles.

---

<sup>72</sup> As noted above about one quarter of employees are assigned to the highest manager quartile in each year and about one quarter are assigned to manager quartiles 4 and 5.

<sup>73</sup> Many different employee performance metrics are used by individual managers in making compensation and quartile assignments. However, only 360 review scores and production are available to me as quantifiable measures of performance and even these metrics are not available for all employees for all years.

<b>Table 1</b> Gender Differences in Placement into Highest Quartile Securities, IBD, and IMD Divisions				
<b>Division</b>	<b>Corporate Title</b>	<b>Years Analyzed</b>	<b>Female-Male Difference in</b> <b>Placement Rate into Top Quartile</b> <b>(in Percentage Points)</b>	<b>Standard Deviations</b>
Securities	Associates	2007 – 2011	0.19%	0.08
Securities	Vice Presidents	2007 – 2011	-0.32%	-0.19
IBD	Junior Banker	2005 – 2011	0.19%	0.08
IBD	Senior Banker	2005 – 2011	-5.16%	-1.23
IMD	Associates	2008 – 2011	-1.23%	-0.28
IMD	Vice Presidents	2008 – 2011	-0.08%	-0.03

Table 2 presents analogous results for the placement of employees into the lowest manager quartiles. A positive gender difference means that women are more likely to be placed above the two lowest manager quartiles, relative to men. The results show insignificant gender differences in the manager quartile assignment process in *all* Divisions for *all* corporate titles.

<b>Table 2</b> Gender Differences in Placement Above Lowest Quartiles Securities, IBD, and IMD Divisions				
<b>Division</b>	<b>Corporate Title</b>	<b>Years Analyzed</b>	<b>Female-Male Difference in</b> <b>Placement Rate above 4th/5th</b> <b>Quartile (in Percentage Points)</b>	<b>Standard</b> <b>Deviations</b>
Securities	Associates	2007 – 2011	2.27%	1.28
Securities	Vice Presidents	2007 – 2011	0.97%	0.66
IBD	Junior Banker	2005 – 2011	0.29%	0.16
IBD	Senior Banker	2005 – 2011	-3.29%	-0.96
IMD	Associates	2008 – 2011	6.50%	1.62
IMD	Vice Presidents	2008 – 2011	-2.08%	-0.91

To summarize the results above, six gender differences are favorable to women and six are favorable to men. I find no significant systematic relationship between gender and manager quartile assignments. The statistical evidence contradicts Plaintiffs' allegations of a common adverse policy or practice, or biased decision-making through the manager quartile process. If Plaintiffs' allegations were true, one would expect to observe common and significant gender differences in manager quartile assignments adverse to women in all corporate titles and all Divisions. Instead I find that gender differences in manager quartile assignments are relatively small and insignificantly different from zero when comparing similarly productive men and women.

### **XIII. GOLDMAN SACHS DATA SHOW WIDE VARIATION IN PAY**

The statistical technique I have used to analyze pay differences is called regression analysis. Regression analysis is an attempt to answer statistically the question of what gender differences remain when comparing individuals who are considered similarly situated under the particular model being used — that is, controlling for the factors that are in the model. This method allows for a measure of statistical “control” for various quantifiable and observable variables.

While simple in theory, the application of this technique is much less precise than the mathematical precision of the calculations would suggest,<sup>74</sup> and its reliability can vary depending on the complexity of the process being analyzed. The highly individualized nature of pay decisions and the extremely wide dispersion in pay from top to bottom among Associates and Vice Presidents mean that regression models may not be as reliable for understanding patterns of pay differentials at Goldman Sachs as they might be at a different employer in a different industry.

I first consider pay dispersion only among men before addressing the issues posed by Plaintiffs' allegations in order to understand the variability of pay in this population without any possible gender effect. This is a way to understand the complexity of the Goldman Sachs data and the ability of regression models to predict pay accurately in the studied population. The table below presents the remarkable fact that there has been more dispersion in total compensation among male Associates and Vice Presidents at Goldman Sachs than among all men employed in full-time (at least 35 hours per week) and full-year jobs (at least 50 weeks) in the entire United States from 2002 to 2011. Most companies do not have employees filling roles of such disparate value as Goldman Sachs, and this phenomenon drives the highly individualized nature of compensation decision-making at the Firm and provides insight on the limitations of regression analysis in this instance. The measure of dispersion used in the table is the ratio of annual compensation at the 90<sup>th</sup> percentile of the distribution to compensation at the 10<sup>th</sup> percentile.<sup>75</sup>

---

<sup>74</sup> Unlike most union pay-setting processes or piecework compensation, salary setting in most occupations, and virtually all service sector occupations, includes subjective assessments of performance and the context in which that performance took place. For example, achieving satisfactory results in an impossible situation may merit higher praise than excellent performance in a smoothly running operation. Weighing these issues in complex tasks requires subjective evaluations, which cannot readily be modeled through regression analysis.

<sup>75</sup> This measure of dispersion in compensation or wages is common in economic studies of inequality, e.g. see Chinhui Juhn, Kevin M. Murphy, and Brooks Pierce "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, June 1993, vol 101, no. 3, pp. 410-442.

Table 3			
Comparison of Internal and External Compensation Distributions, 2003-2011			
	10th	90 <sup>th</sup>	90-10
	Percentile	Percentile	Differential
Full Time/Full Year Male Pay in External Labor Market			
(Source-Current Population Survey)	\$18,944	\$100,600	5.31
Male PATC at Goldman Sachs			
(Associates & VPs, All Divisions)	\$[REDACTED]	\$[REDACTED]	[REDACTED]

Using this measure, in each year the dispersion in compensation among Goldman Sachs Vice Presidents has been larger than in the general population of men nationwide. This empirical fact is remarkable because pay dispersion in the general population involves comparisons among workers with *wide* differences in their skills, experience, occupations and educational attainment. Men near the top of the overall U.S. earnings distribution make 5 to 6 times more than men near the bottom of the distribution in part because the population includes men with advanced degrees and those who did not complete high school. Earnings dispersion in the general population is wide because it includes comparisons of men in vastly different occupations, including, at the high end of pay, attorneys and physicians, and, at the low end, janitors and dishwashers. Pay dispersion in the general population compares men just out of school at the start of their careers and experienced men at the peak of their careers, as well as employees at different companies in different industries in a variety of locations.

Labor economists who study pay differences use generic factors such as years of labor market experience, educational attainment, occupation and industry to account for some of the pay differences in the general population. However, because Goldman Sachs Vice Presidents have rather similar educational backgrounds and work experience, and are employed in the same broad occupation at the same firm, none of the standard control factors available to labor economists can be used to explain variation in pay in this proposed class. Instead, compensation differences among Goldman Sachs employees depend largely on market conditions and individual circumstances. Although there is more compensation variation to be explained among Goldman Sachs Vice Presidents than in the general population, there are fewer measured factors and numerical characteristics at the statistician's disposal to explain these compensation



differences. This is not because Goldman Sachs does not maintain many forms of personnel data. But the individualized nature of the work being done and the breadth of the compensation range means that statistical analysis will have a hard time explaining all of the variation. Indeed, a regression model will provide less insight or explanation of pay differences among putative class members and men who are Associates and Vice Presidents at Goldman Sachs than in the general workforce.

#### **XIV. MY REGRESSION ANALYSIS PRESENTS A FULLER AND MORE ACCURATE ANALYSIS OF COMPENSATION**

In a statistical analysis of employment practices, the analysis compares group outcomes. Here, we compare female to male promotion rates, pay and other employment outcomes. Statistical analysis allows us to answer the question of whether the differences we see in the data are so large that chance, or random outcomes, are unlikely to have been the cause of the differences.<sup>76</sup> These differences are most often measured in terms of statistical significance. “Statistical significance” is not an absolute concept but is, rather, a matter of convention. When we see an event that is unusual, we compare it to our experience to judge just how rare the event was – lightning striking twice, and so on. In statistics, there are mathematical concepts for judging how rare an event is given the range of random outcomes that occur in real data. Statistics can quantify how unusual certain outcomes are. In particular, statistics can say how often we would expect to see an extreme event if only chance were guiding the generation of the outcomes. In statistics, the more standard deviations that separate the statistical results for the protected group and the majority group the less likely that the difference was generated by chance.

There is a continuum of possible cut-off points for judging when a difference is so large that such random outcomes, or “noise,” must be rejected as the most likely explanation. The Supreme Court in *Hazelwood School District v. United States*, 433 U.S. 299 (1977), gave some guidance by suggesting that the difference should exceed “two or three standard deviations.” Social scientists doing empirical work have generally adopted the “two standard deviation” rule-of-thumb. I have also adopted that convention in my report when speaking of “statistical significance.”

##### **A. “Similarly Situated” Employees**

In studying pay I have attempted to compare men and women who are “similarly situated.” This is accomplished by the use of statistical methods that attempt to “control for” or “hold constant” various factors that may affect pay, apart from gender, for example, Business Unit and manager quartile. It is important to base analyses of group differences on

---

<sup>76</sup> Random is a term of art in this context. There is no implication in its use that the decisions at issue are actually random when made by the decision-makers at the employer. The term refers to the inability of an outside observer to explain differences between the compensation of different employees with the model.

comparisons between similarly-situated employees, and this basic principle is well-established in both economics and the law.

For example, the U.S. Equal Employment Opportunity Commission's (EEOC's) Compliance Manual defines "[s]imilarly situated employees [as] those who would be expected to receive the same compensation because of the similarity of their jobs and other objective factors." (10-III, A. 1.) Further, the Manual directs EEOC investigators: "When beginning an investigation for compensation discrimination, it is important to acquire information about the respondent's general system for compensating its employees. It will be useful to identify employees similarly situated to the charging party for purposes of comparing their compensation." (10-III (A,1.,a). In this case, I rely upon the major features of the compensation process at Goldman Sachs in order to analyze together people who are relatively similarly situated. In particular, I take similarly situated employees to be those subject to similar business conditions, engaged in similar work and achieving similar performance.

#### **B. Similar Business Conditions**

As described earlier, Goldman Sachs budgets compensation in its annual cycle, first at the level of the entire Firm, then to Divisions and finally to Business Units and departments or desks within those Business Units. The Business Units' allocation of aggregate compensation changes (up or down) is based on the profitability of all three levels; the managers who make the initial recommendations for compensation do so operating within the budget of the Business Unit. Therefore, my statistical analysis considers employees to be similarly situated if, as one factor, they are in the same Business Unit. Failure to analyze Business Unit in a compensation model would be proper only if two employees should, on average, have the same compensation, holding other factors constant, even if one employee was in a highly profitable Business Unit while the other was in an unprofitable Business Unit, and that is not the case.

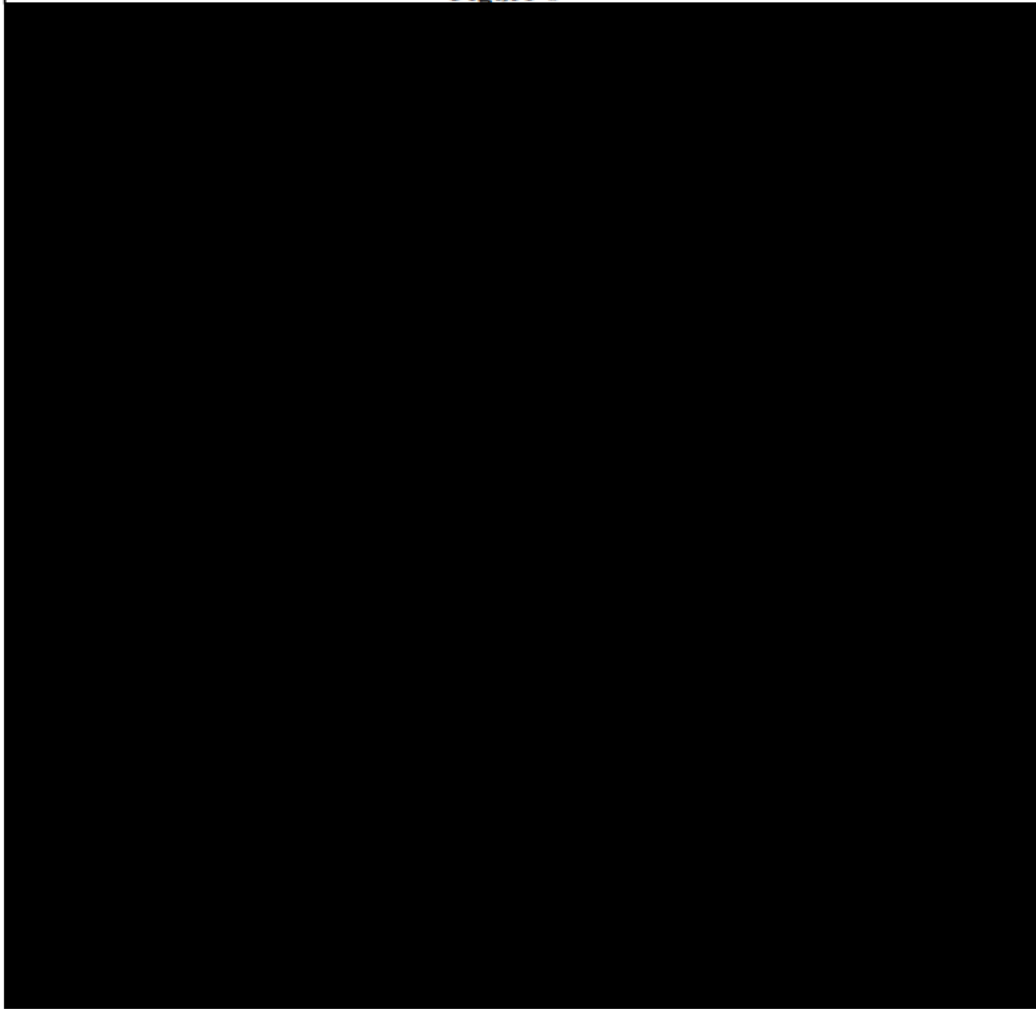
The [REDACTED] is also important. For example, in the largest Division, Securities, [REDACTED] is a key metric as described in the 2011 Securities Division Compensation Guidelines.<sup>77</sup> (See Figure 6). [REDACTED]

[REDACTED] Statistical models comparing employees at this level, while they would most accurately capture the business conditions affecting employees' compensation, would result in few comparisons of men and women. Thus, here, as is often the case, regression models (which rely on quantifiable data) require compromises with reality in order to study a meaningfully sized group.

---

<sup>77</sup> GS0113896: 2011 Securities Division Compensation Guidelines.

**Figure 6**



**C. Similar Work**

Goldman Sachs does not identify an employee's "job." Employees working in a Business Unit are distinguished in the data by "function" and also by "department." Department indicates not only the organizational unit within which an employee works, but also the product or client group with which the employee deals.

Functions are a very broad grouping of activities: Sales, Trading, Strategies, Securities Services, Finance, Sales-Trading, Principal Investing, Specialist, Electronic Services and finally a "Miscellaneous" category. For example, "Sales" encompasses marketing to corporate clients, municipalities, and hedge funds as well as other investment banks. Strategies ("Strats") includes employees developing complex mathematical models as well as those sitting at a "desk" and assisting traders in implementing financial strategies. As further described on the Goldman Sachs website,

The professionals in this group develop quantitative and technical solutions in partnership with teams throughout Goldman Sachs. Their backgrounds include science, mathematics or engineering as well as economics and finance. In addition to traditional quantitative finance activities, such as pricing and risk management, Strats are applying their skills to areas that include corporate finance, regulatory reform, business intelligence and automated trading.<sup>78</sup>

As another example, the “Finance” function includes those who assist corporations in managing risk with derivatives, assisting with equity or debt financing, bank loans, leveraged buyouts or restructurings. The Goldman Sachs web site at “What We Do” gives a broad overview of the kinds of functions performed; even these brief descriptions are clearly much more detailed than is indicated in the computerized human resource data under the variable “Function.”<sup>79</sup>

Finally, unlike Dr. Farber, I analyze employees who are Associates separately from Vice Presidents.<sup>80</sup> Managers within Business Units of Goldman Sachs separately determine manager quartiles for employees at different levels and, sometimes within those levels by their seniority.<sup>81</sup> This corresponds to their different skill and experience level.

#### **D. Similar Performance**

Finally, I compare people who are similarly situated with respect to their performance. Manager quartile and 360 reviews provide two measures of performance at Goldman Sachs. Additionally, performance as evidenced by individual production is taken into account in setting compensation—both current and prior year’s revenue production. The Firm rewards “proven track records.” For example, the 2011 Securities Division Compensation Guidelines provides this summary:

In making compensation determinations, compensation managers should take into account both current year financial performance and financial performance over time. The Firm seeks to reward proven track records, and a review of revenue and RNOVs attribution over multiple years allows compensation determinations to reflect financial performance under a variety of market conditions and to better take into account low-probability risks and other risks that are identified over time. The Firm’s Compensation Principles emphasize that “[e]ffective compensation practices should evaluate performance on a multi-year basis.”

Each Division has its own guidelines, which include key metrics that differ substantially from Division to Division.

See, for example, Figure 7 below.<sup>82</sup>

---

<sup>78</sup> Goldman Sachs, *Securities – Strats*, <http://www.goldmansachs.com/what-we-do/securities/products-and-business-groups/business-groups/strats.html>.

<sup>79</sup> Goldman Sachs, *What We Do*, <http://www.goldmansachs.com/what-we-do/>.

<sup>80</sup> In IBD, I separate Junior Bankers from Senior Bankers. In Dr. Farber’s Rebuttal Report he, belatedly, performed separate analyses of Associates and Vice Presidents.

<sup>81</sup> Donovan Deposition Exhibit 132; GS109288-9; GS0115071-2.

<sup>82</sup> Larson Deposition Exhibit 141, 2011 Investment Banking Division Compensation Guidelines (GS0113426); GS0113446; GS0113917-9.

**Figure 7****E. Many Unmeasurable Factors and Approximate Values Affect Variation in Compensation**

The factors described above, of business conditions, role, and performance, are not as precisely quantified by the available data as a statistician would like. I had to work with values that I knew did not capture all relevant differences between employees. For example, the need to analyze Business Unit instead of desk or department leads to grouping different industry or product groups together (not to mention groups of employees who were paid out of different desk or department budgets). That in turn reduces the ability of the model to explain pay with precision, even for men alone. Similarly, I am limited in my ability to study people who perform the same job. The wide variety of roles and the lack of any code for them means that I must use Division function and corporate title to capture some, but by no means all, of the differences between employee activities.

Last, while it would be more accurate if I could account for an employee's long term history of performance and production, I am once again limited by the data. I control for current dollar value of production and also the prior year's production. However, for many Business Units and job functions production data is not available or was not captured in earlier years. Also, Individual Strats do not have financial production. All Strats look the same to an outside observer of the data even though their performance may vary based on whether or not they are attached to a desk that performs unusually well.

This challenge is demonstrated by consideration of another statistic that is a byproduct of the regression analysis. The Root Mean Square Error ("RMSE") of the regression model measures the variability of PATC for individuals with the

Confidential Pursuant to Court Order and Data Handling Agreement

same predicted pay from the regression model. It is another measure of the model's precision. For example, if we took one person with a given value for year, Business Unit, location, function, education, guarantee, 360 score, manager quartile for this and the prior year, and production for this and the prior year, we might hope that this person's PATCs might be estimated fairly accurately. In fact, the "error" or range associated with the predicted pay of these employees is enormous at Goldman Sachs. The RMSE gives a measure of this imprecision. I calculated the interval within which actual PATC may fall when the regression model predicts a PATC of \$ [REDACTED].<sup>83</sup> I used the results for Vice Presidents in Securities shown in Table 4. The high end of the range is \$ [REDACTED] while the low end is \$ [REDACTED].

This error or imprecision represents ignorance about many pay setting elements by outside observers of data—like the statistical experts in this litigation. As discussed earlier, the compensation process at Goldman Sachs incorporates many factors which are not maintained in computerized data or are not easily quantifiable. The key assumption in regression analysis is that all of those factors that are not observed have no correlation with gender. The larger the unexplained variation in pay, even with the most sophisticated model possible, the more suspect is the assumption that all unobserved differences are in fact gender-neutral.

Moreover, the factors described above do not tell the whole story when it comes to drivers of compensation differences. As described in the Divisional Compensation Guidelines, individual allocation of compensation within the Business Unit is based on many other factors, most of which are not available in the computerized data that I rely upon:

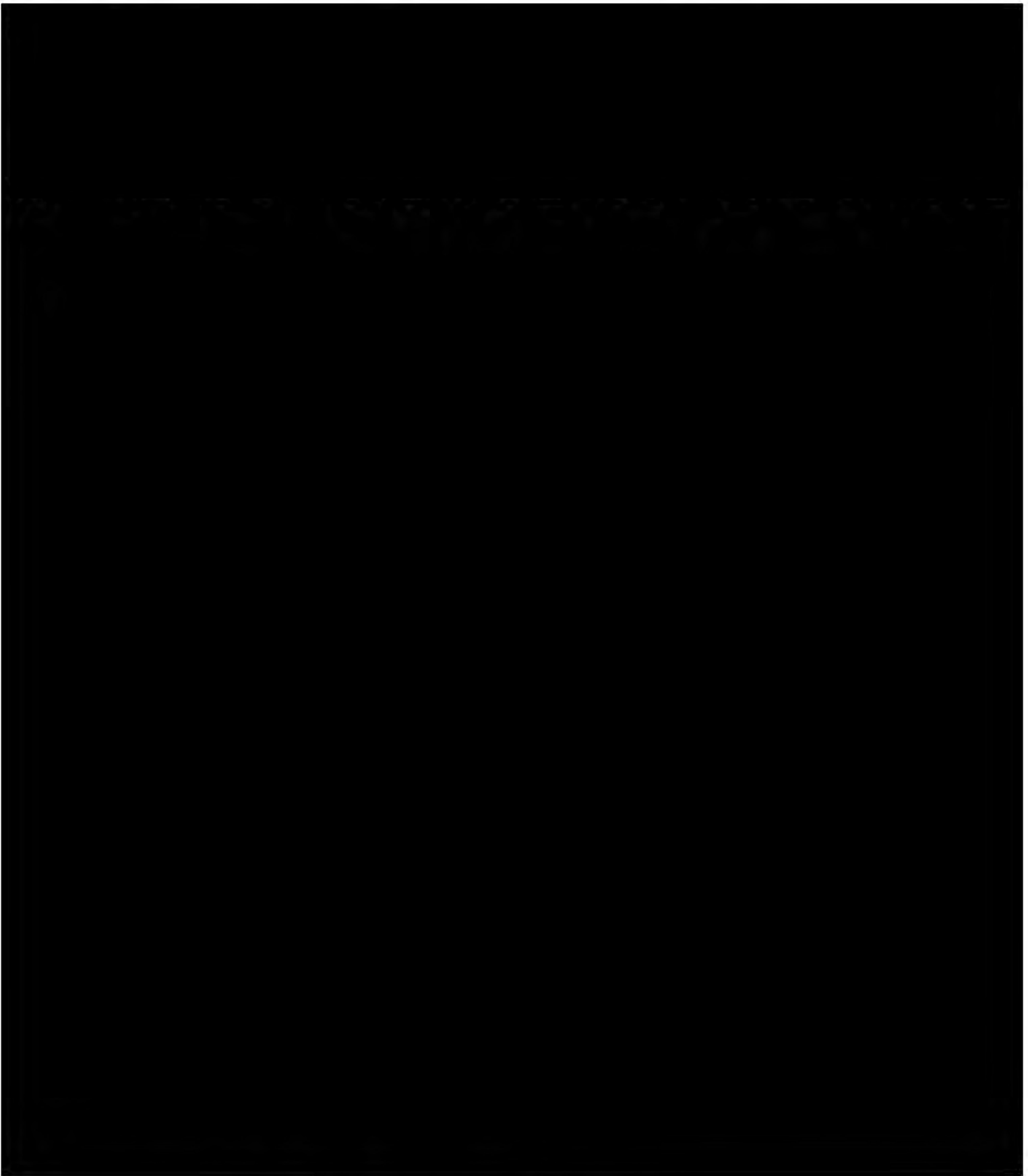
[REDACTED]  
[REDACTED] (See Figures 6 and 7). These legitimate factors influence compensation decisions but cannot play a part in my equations.

I prepared a set of what I call "matched" employees, all of them Vice Presidents, who shared the same values for factors included in my statistical model: Business Unit, department, job function, tenure at Goldman Sachs, manager quartile, and whether they were hired as a lateral. Among those who matched I looked for pairs with very different compensation in the same year, regardless of their genders. I then asked the managers who supervised these individuals to explain, if they knew, why the compensation of apparently identical individuals was so different.<sup>84</sup> This is an approach I have used to try to identify important pay factors that I may have overlooked or factors that are simply not recorded in data.

---

<sup>83</sup> The interval I calculate is a 95% confidence interval for PATC.

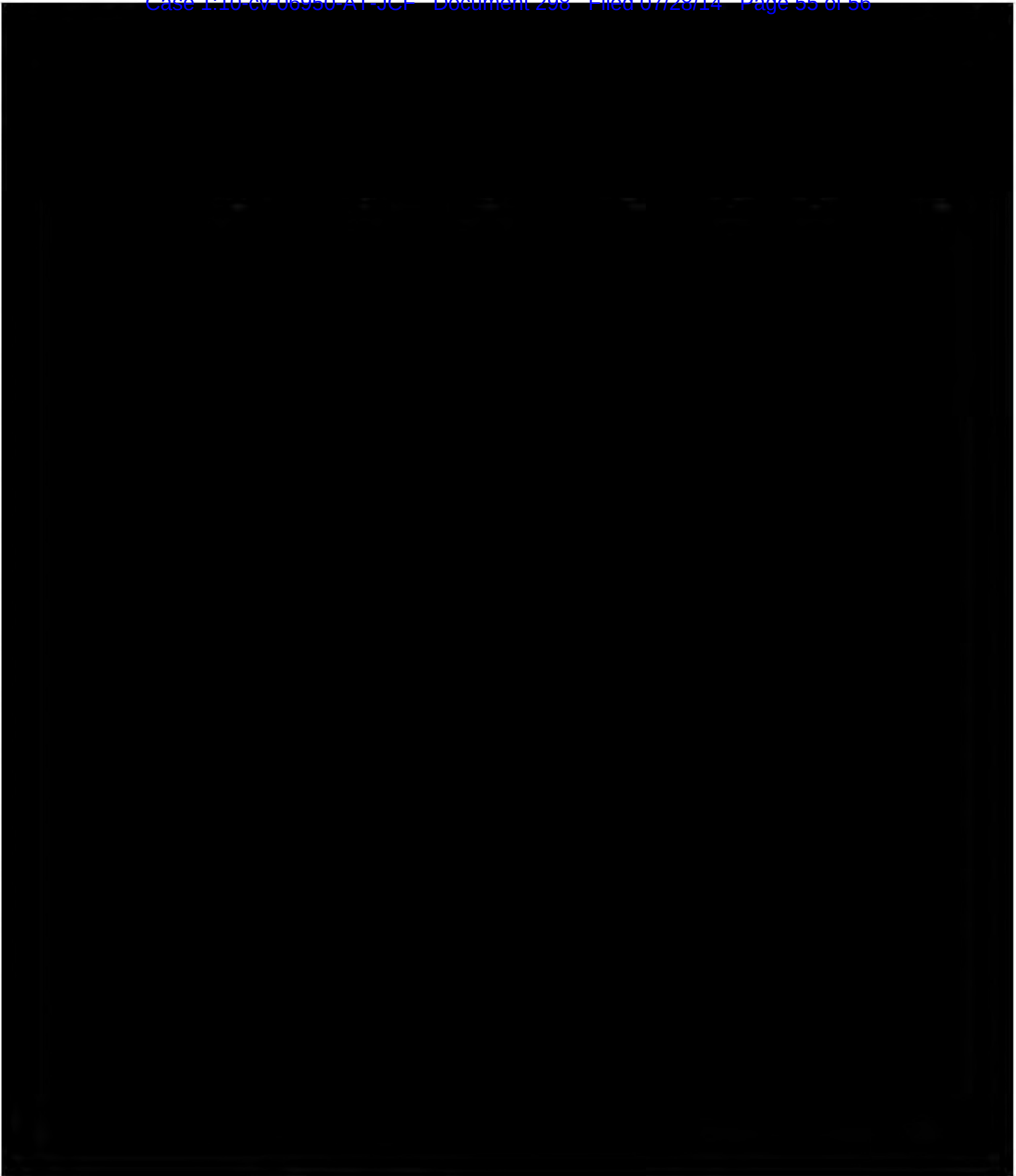
<sup>84</sup> In some cases there were more than two "pairs" because three people matched on these characteristics.



---

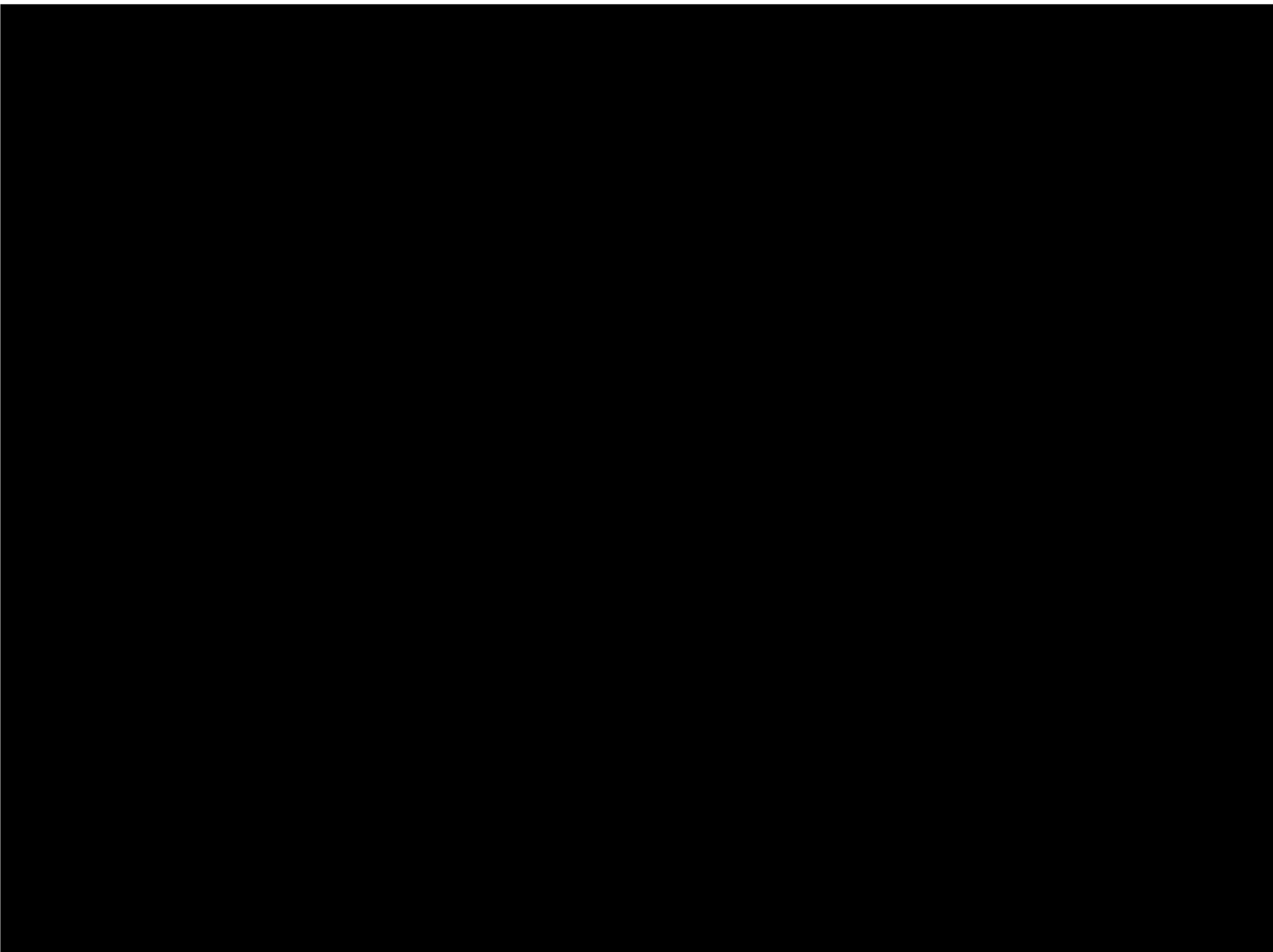
[Redacted]





---

[Redacted text block]



In each of these comparisons, the higher compensated person worked in a role that was materially more valuable than the lower compensated person: acting as a team lead or having significantly greater responsibility and financial contribution to the Firm. These differences are not reflected in the computerized data available from Goldman Sachs. The examples show that significant compensation differences remain even after “controlling for” the factors that are available for a regression analysis. The regression model is assuming that these factors are, on average, the same for men and for women, even though there is no way to test that assumption with the data. It requires discussion with decision-makers, who in these instances provided concrete examples of the differences in skill sets, contribution to Firm revenue, and external market value that cannot be captured in computerized data.

**F. My Regression Models Better Reflect These Concepts Than Dr. Farber’s**

In order to incorporate the concepts discussed above, I include the following factors in my regression analyses, in addition to gender:

- Fiscal Year